POPULATION GENETIC PATTERNS IN SISTER SPECIES PAIRS SHARING A SINGLE SPECIATION EVENT: A STUDY OF TWO SPECIES PAIRS OF FRESHWATER FISHES

By

Anna Carlson

Submitted to the Department of Ecology and Evolutionary Biology and the
Faculty of the Graduate School of the University of Kansas
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

_____
Chairperson

_____
Chairperson

Committee members _____

_____

_____

Date defended: _____

The Dissertation Committee for Anna Carlson certifies
That this is the approved version of the following dissertation:

POPULATION GENETIC PATTERNS IN SISTER SPECIES PAIRS SHARING A SINGLE SPECIATION EVENT: A STUDY OF TWO SPECIES PAIRS OF FRESHWATER FISHES

Committee:

_____
Chairperson

_____
Chairperson

_____

_____

_____

Date approved: _____

Table of contents

Chapter 1: Introduction

What happens to microevolutionary parameters (allele frequencies and distributions) during macroevolutionary changes (speciation)? This question has rarely been addressed in fishes and is a difficult one to study. Evidence suggests that historical demographic and zoographic factors commonly shape microevolutionary parameters (Avise, 1989), so it might be expected that any microevolutionary changes occurring during the speciation process would soon be masked by later changes. Replicate and recent speciation provides some control for such historical factors, however. This study examines microevolutionary effects of speciation. Specifically, does the observed population genetic structure of daughter species show a pattern indicative of change caused by the speciation process itself? Alternatively, is there no change or change idiosyncratic to each species? The available evidence is the current observed population structure of sister species (daughter species formed in a speciation event). If two sister species have the same population genetic structure, the most parsimonious inference is that the ancestral species also had the same population structure and that this structure was passed down unchanged through the speciation event. If, however, two sister species differ in population genetic structure then all that can be inferred is that at least one, possibly both, of the sisters differs in population structure from the ancestral species—the change in population genetic structure could be due, for example, to a change in breeding habits or a bottleneck that occurred since speciation and cannot be linked to the speciation event. Figure 1.1 illustrates these possibilities with one species pair.

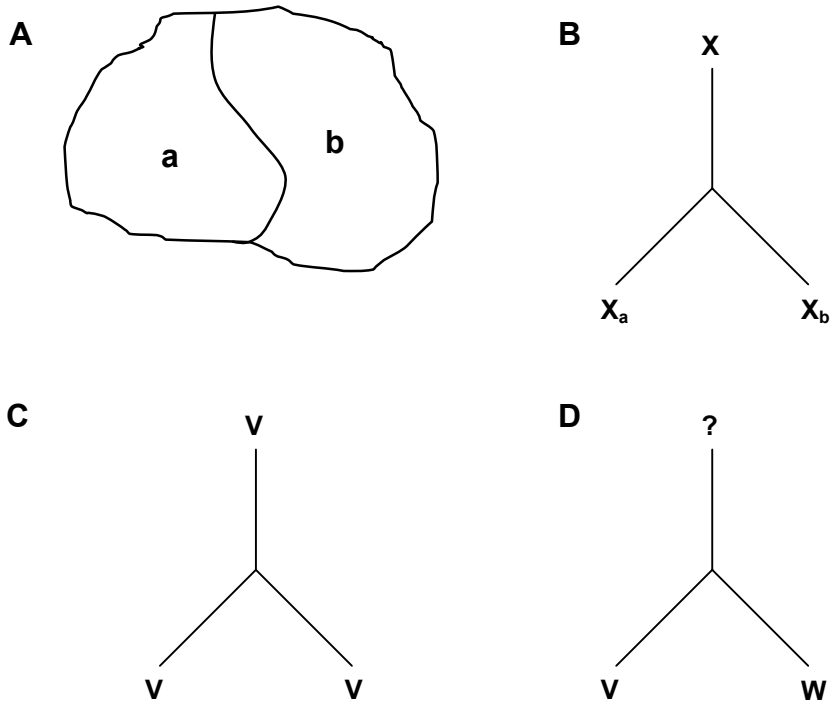


Figure 1.1. Ranges of sister species, and values of a population genetic parameter shown on phylogenetic diagram for ancestor and daughter species. (A) Adjacent ranges of sister species **a** (on the left, or west) and **b** (on the right, or east) as observed in the present. (B) Depiction of history that brought about the situation observed in part A. Let X be a value for a population genetic parameter in an ancestral species at some point in the past (not observable), X_a is the value of that parameter currently observed in daughter species **a** and X_b is the value of that parameter in daughter species **b**. Lines show phylogenetic descent through time from ancestral to daughter species (present time at bottom). (C) If observed values for X_a and X_b are the same (let them equal value V), the most parsimonious inference is that the value for that parameter was also V in the ancestor and the value was passed down unchanged through the speciation process. (D) If the observed values for X_a and X_b are not the same [X_a = V and X_b = W, V ≠ W], then it is unknown whether X_a, X_b, or both changed from the ancestral value, and the ancestral value cannot be inferred. Moreover, there is no indication whether the population genetic values changed primarily during the speciation process or after it was completed.

Change in population genetic structure caused by the speciation process itself can be revealed by examining systems of replicate speciation, where several species

pairs formed in parallel during the same vicariant event. Each pair of sister species is one sampling of the speciation process. Ideally, the pairs would be distantly related, one to another. If the pairs show the same pattern of difference in population genetic structure, it is reasonable to infer that this change was brought about by the shared vicariant event. For example, if each pair shows a lower value for a given parameter in the western sister compared to the eastern sister, a parsimonious explanation is that the speciation event itself, which was shared by all species pairs, brought about a change in the value of the parameter in sisters on one side of the vicariant line. This will be referred to as a "similar pattern of change." Figure 1.2 presents the example of two species pairs; 1.2C illustrates a similar pattern of change. Of course, change caused by speciation is not the only explanation for such an observed pattern but it is the simplest one in the absence of information about other factors common to sisters on the same geographic side of the event. The more sister pairs involved, and the more distantly related and different in natural history they are, the stronger the argument that population genetic change occurred due to the speciation event.

Relative to previous study of the genetics of speciation, the current work links microevolutionary and macroevolutionary study in a novel way: Does history of speciation (macroevolution) have predictive power for patterns of alleles in populations (microevolution)? The focus of the majority of work combining genetics and speciation is on the cause of reproductive isolation (see Coyne and Orr, 1998 and 2004, for reviews of literature on speciation). With this aim, many authors have related strength of the isolating barrier between species to genetic distance (Coyne

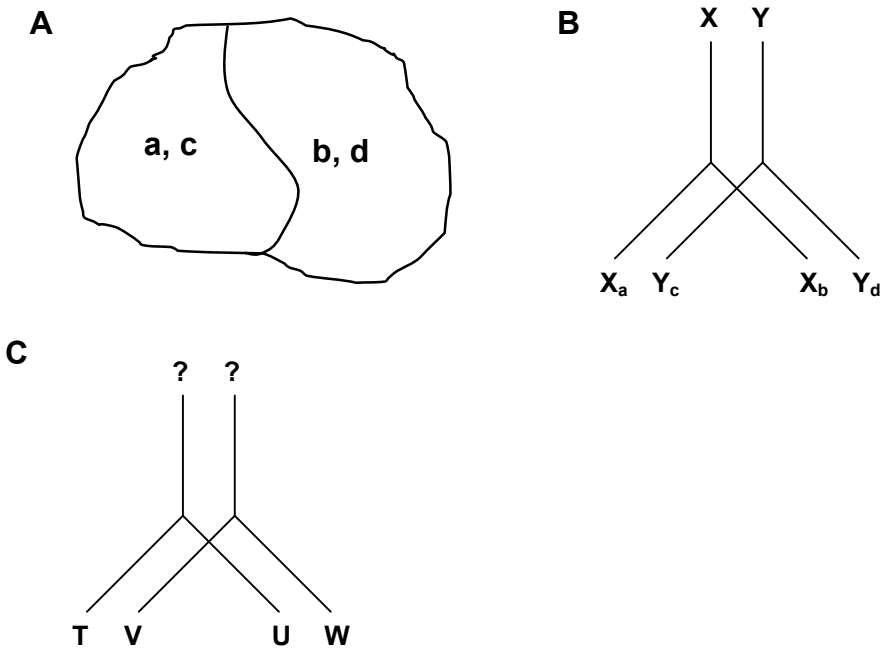


Figure 1.2. Ranges of two pairs of sister species, and values of a population genetic parameter shown on phylogenetic diagrams. (A) Range shared by two pairs of sister species; species **a** and **b** comprise a sister species pair, as do species **c** and **d**. The two pairs may be very distantly related to each other. The species pairs share a geographic inter-species dividing line and are distributed such that species **a** and **c** are sympatric (on the left, or west) and **b** and **d** are sympatric (on the right, or east). (B) Depiction of history that brought about the situation observed in part A. Let X be the value for a population genetic parameter in an ancestral species and Y be the value in another ancestral species at some point in the past (not observable), X_a and X_b are present values of that parameter observed in daughter species **a** and **b**, respectively, and Y_c and Y_d are the values of that parameter in daughter species **c** and **d**, respectively. (C) Let $X_a = T$ and $X_b = U$ such that $T < U$ and let $Y_c = V$ and $Y_d = W$ such that $V < W$. The ancestral value for either species pair cannot be inferred. The two phylogenetically independent species pairs show the same pattern of parameter change, however (the western species has a smaller value compared to the eastern species). Such a pattern is referred to as a "similar pattern of change" in the text. A parsimonious explanation for this is that the parameter was affected in the same way in each species during their shared speciation event.

and Orr, 1989 and 1997; Presgraves, 2002; Price and Bouvier, 2002; Bolnick and Near, 2005; see Lijtmaer *et al.,* 2003, for comparison of rates in different taxa), compared strength of behavioral and intrinsic isolating mechanisms between species (Sasa *et al.,* 1998; Mendelson, 2003; Russell 2003), and examined effects of sexual selection on speciation (Barraclough *et al.,* 1995; Mitra *et al.,* 1996; Arnqvist *et al.,* 2000). Much work has also focused on whether drift or selection is primarily responsible for genetic divergence during speciation (Coyne and Orr, 2004, chapter 11).

Studies examining population genetic outcome of speciation are rare, however. Many vicariant events have been documented, caused by geographic phenomena such as mountain building episodes (*e.g.,* Ruzzante *et al.,* 2006) continental drift (*e.g.,* Iba and Sano, 2007), and hydrogeographic changes such as river canyon deepening (*e.g.,* Murphy *et al.,* 2006), but population genetic analyses of species pairs formed in these events are few and none have been found comparing population genetic parameters among species pairs. For example, formation of the Panamanian Isthmus caused a well-known vicariant event affecting many marine organisms including fishes. A survey of literature regarding this event revealed eleven articles concerning geminate species pairs of fishes formed in this event (Vawter *et al.,* 1980; Graves *et al.,* 1983; Foster 1989; Quinteiro *et al.,* 2000; Bowen *et al.,* 2001; Grant and Leslie, 2001; Muss *et al.,* 2001; Wellington and Robertson, 2001; Bellwood *et al.,* 2004; Bernardi *et al.,* 2004; Craig *et al.,* 2004), of which only one (Muss *et al.,* 2001) calculated any population genetic parameter, and then in only

one sister species pair. Literature searches revealed no articles comparing population genetic outcome of speciation among geminate species pairs. The current study appears to be unique in doing this.

It is instructive to look at the components of the speciation process in light of how they may affect microevolutionary processes. While peripheral isolates, parapatric speciation, or even sympatric speciation may have recognizable effects on the allele frequencies and distributions passed from ancestral to daughter species, only vicariant speciation (model I allopatric speciation in Wiley, 1981) will be addressed here, since it is most likely to bring about replicate speciation. Vicariant speciation involves a geographic barrier separating a species range into two large, genetically isolated subregions. Selection and genetic drift (most likely in combination) then cause genetic divergence of the populations in each subregion until two distinct species are formed. Important components of this speciation process that could affect microevolutionary processes include the genetic isolation, itself, and the geographic characteristics of the region each isolated population inhabits, as well as forces of selection on the separated populations. Instances of vicariant speciation are inseparable from the geography in which they occur, so the geographic particulars of the location in which each population finds itself can be considered part of the dynamics of the speciation event. Since in vicariant the two separated populations are both large, population size reduction, per se, is unlikely to affect microevolutionary processes dramatically, although if one population is smaller than the other a reduction in genetic diversity may be seen due to increased drift. In conjunction with

geography, however, genetic isolation of the two populations may have a distinct effect on microevolutionary processes. One of the isolated subregions may be more fragmented ecologically and more restrictive to internal migration. Once genetically separated from the other subregion, genetic subdivision within this region may increase. Selection can also affect microevolutionary processes. Selective pressure in one subregion could increase frequency of some alleles and accelerate loss of others. It might also have a genetically homogenizing effect within the subregion if certain genotypes are favored over the extent of the subregion. Other (perhaps unknown) forces may affect microevolutionary change during the process of speciation; these are simply some examples showing how such change could conceivably happen.

As previously discussed, microevolutionary change during speciation cannot be detected by examining one species pair in isolation. One of the predicted characteristics of vicariant speciation, however, is that speciation will be replicated in phylogenetically disparate groups across the same geographic line of genetic separation (vicariant line) because a geographic phenomenon (mountain building, continental separation, stream capture, etc.) causes genetic isolation in many unrelated taxa. Conceivably, the dynamics of a vicariant event may affect the microevolutionary characteristics of unrelated taxa in similar ways, bringing about a similar pattern of change in geminate species. For instance, if one subregion is more fragmented and limits migration for one daughter species, the causes of the fragmentation may affect other taxa as well. The result would be a replicate

speciation event in which phylogenetically disparate sister species pairs all have higher genetic subdivision in the sister found in the more fragmented subregion. Another possibility is that environmental factors causing selection pressure on one daughter species may also exert pressure on daughters in other sister pairs. In this case, sisters in one subregion may have lower genetic diversity than those in the other subregion. In each case, a similar pattern of change would be observed. It is also possible that such a pattern may be seen due to demographic or zoographic changes affecting only species in one subregion subsequent to speciation, but this is less likely in the absence of known historical events thought to have such an effect, especially in cases of recent speciation events.

The current study uses two species pairs of freshwater fishes to address population genetic effects of speciation. The pairs, one of sand darters and one of starhead topminnows, live in streams in the Gulf Coastal region of the southeastern United States and are both thought to have speciated during a vicariant event that occurred within the last 2–4 million years (Price and Whetstone, 1977; Wiley, 1977; Wiley and Mayden, 1985). The sand darter pair comprises *Ammocrypta beanii,* the naked sand darter, and *A. bifascia,* the Florida sand darter; the starhead topminnow pair comprises *Fundulus nottii,* the southern starhead topminnow, and *F. escambiae,* the russetfin or eastern starhead topminnow. The two pairs are distantly related to each other taxonomically and are distinct in many aspects of natural history, though sister species within each pair share similar natural history. Details of the taxonomy and natural history of each species pair is given in the chapters about those pairs

(Chapter 2 for *Ammocrypta* and Chapter 3 for *Fundulus),* and a comparison of the pairs' natural history is given in Chapter 4.

Two types of molecular data were collected and analyzed for each species pair: DNA sequence data for the mitochondrial cytochrome *b* gene, and nuclear AFLP (Amplified Fragment Length Polymorphism) fingerprints. The cytochrome *b* sequence was used in a phylogenetic analysis to determine group affinities of individuals from different locations in the analysis. This was necessary because question arose as to the phylogenetic affinity of *Ammocrypta* populations from the Mobile Bay Drainage (see Chapter 2). The sequence data were also used to generate an estimate of between-species divergence in each pair. AFLP fingerprints were used to assess the population genetic parameters of polymorphism, average heterozygosity, and F_{st}, as well as to derive an estimate of between-species divergence independent of that derived from the cytochrome *b* data. AFLP data were also used to examine possible isolation by distance among the populations within each species. For methods, see Chapters 2 and 3.

Population genetic structure of each species, as well as comparison of species within each pair, and discussion of possible inferences about the population genetic structure of the ancestral species are given in the chapters on the individual species pairs (Chapter 2 for *Ammocrypta* and Chapter 3 for *Fundulus)*. Chapter 4 compares population structure between the two pairs, and discusses possible inferences about the speciation event itself. General findings are that population subdivision (measured by θ^B, an analog of Wright's F_{st}) is much higher in the *Fundulus* species

than in the *Ammocrypta* species and appears to have been so in the ancestor, as well. The speciation process appears to have had little or no effect on population subdivision and average genetic distance between populations in these two species pairs. However, genetic variability (as measured by gene diversity) does show a pattern consistent with change caused by the speciation event. This work outlines a new process for the study of the population genetic consequences of speciation, although inferences about this vicariant event are of necessity tentative since only two species pairs were examined. Future work could reveal whether the results found here characterize other species pairs from this vicariant event, or even the speciation process on a wider scale.

Chapter 2: Examination of phylogenetic relationships and population genetic characteristics in *Ammocrypta bifascia* and *A. beanii*

The goal of this part of my work is to determine phylogenetic affinity of the Mobile Bay Drainage populations of *Ammocrypta beanii* and examine population genetic characteristics within the sister species pair *Ammocrypta beanii* and *A. bifascia,* which speciated within the last 4 million years (Wiley and Hagen, 1997). Characterization of population genetic structure of closely related species enables inferences to be made about the genetic structure of the ancestor. This is the first assessment of the population genetic structure of any species in the genus *Ammocrypta.*

Introduction

Rationale

Do speciation events leave a recognizable population genetic pattern in the species formed? To begin addressing this question, a sister species pair, *Ammocrypta beanii* and *A. bifascia,* was identified that has three key characteristics: (1) the date of the speciation event was approximately known and was recent enough to expect a population genetic pattern caused by speciation to still be evident, (2) evidence suggested there is no third species, extinct or extant, that is more closely related to either species of this pair than they are to each other (Wiley and Hagen, 1997; Williams, 1975; Smith, 1981), and (3) the occurrence of each species within several river drainages allows clear demarcation of population subdivisions for population

structure analysis. In addition, specimens were available in the KU Natural History Museum tissue collection.

Sequence information from the mitochondrial cytochrome *b* gene was chosen for the phylogenetic analysis because this gene is straightforward to sequence, has an appropriate amount of variation for the question at hand, and outgroup sequences were available on GenBank. AFLP, or Amplified Fragment Length Polymorphism, was used to study population genetic parameters because it yields a large number of genetic markers, has been shown to be highly reliable, surveys the entire genome, and requires no prior knowledge of genome sequences.

Study organism

Ammocrypta beanii and *A. bifascia* are two species of sand darter that inhabit streams in the northern Gulf Coast region of the United States. They lack scales over most of the body and mature at 4–6 cm standard length, with *A. bifascia* generally larger than *A. beanii*. Williams (1975) described *A. bifascia*, and differentiated it morphologically from *A. beanii* based on fin banding pattern and breeding tubercle distribution. *A. beanii* and *A. bifascia* have been found to be sister species in all phylogenetic studies in which they have been included (Shaw *et al.*, 1999; Wiley and Hagen, 1997; Near *et al.*, 2000). Near *et al.* (2000) found ten unique nucleotide and five unique morphological synapomorphies supporting this sister relationship. In addition, these species occupy adjacent, non-overlapping ranges (Lee, 1980), a distribution which suggests that no intermediate species have evolved and subsequently become extinct (Smith, 1981). If the population genetic outcome of a

speciation event is to be assessed, "true" sister species with no extinct species between them are required. As will be discussed later, results of this study indicate that *A. beanii* specimens from the Mobile Bay Drainage (MOB) are genetically distinct from other *A. beanii,* and may form a separate group or even group more closely with the *A. bifascia.* The possible new MOB species has a distribution concordant with that of at least 40 other species of fishes in the area, making it fit well with the biogeography of the area (Swift *et al.,* 1986). For the purposes of this paper, *A. beanii* from drainages other than the MOB will be referred to as "western *A. beanii."*

Methods and materials

Experimental overview

For this project, two different types of molecular data were used: cytochrome *b* gene sequence, and AFLP (or Amplified Fragment Length Polymorphism) fingerprints. The sequence data were used to resolve the phylogenetic affinity of the MOB *A. beanii,* and the AFLP fingerprints were used for population genetic analysis. The AFLP data also provide some information, such as relative distances between groups and group diagnostic loci, that is relevant to phylogenetic questions.

Samples

All samples were collected in 1987–1989 by E. O. Wiley and are housed in the University of Kansas Natural History Museum tissue collection (see Appendix 1). They were frozen in liquid nitrogen when collected and transferred to –80º C for long-term storage. Samples were chosen from as wide a geographic distribution as

possible from within each species' range. For the AFLP analysis, an average of 15 specimens were used from each population (range: 5–34; figure 2.1). All available individuals from the Mobile Bay Drainage were included. These consisted of six individuals taken from the Tombigbee River and 15 from the Alabama River. See figure 2.1 for the location of drainages and sample sites, and for numbers sampled from each site. Excluding the two MOB populations, seven populations were sampled for *A. beanii*, representing four drainages and consisting of 102 individuals.

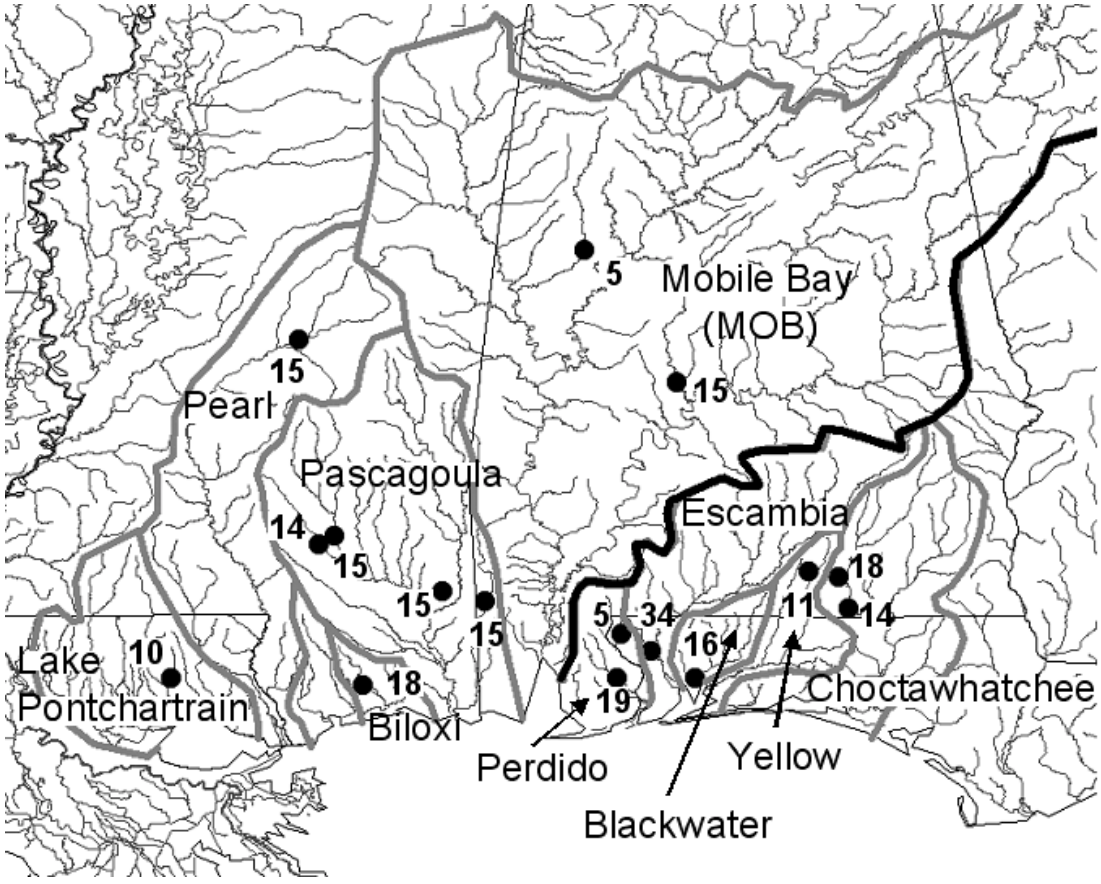


Figure 2.1. Drainages represented by samples used in the AFLP analysis. Grey lines show drainage boundaries. Number of specimens included is shown near each collection site. The black line separates the range of *Ammocrypta beanii* (west) from that of *A. bifascia* (east).

Seven populations representing five drainages (116 individuals) were sampled for *A. bifascia*. For each of these three groups, at least one drainage has more than one population sampled so that genetic structure could be estimated within drainages as well as among them. For the phylogenetic analysis, 19 individuals were sequenced: five *A. beanii* from two locations within the MOB, nine *A. beanii* from three other drainages within their range, and four *A. bifascia* from three drainages (figure 2.2).

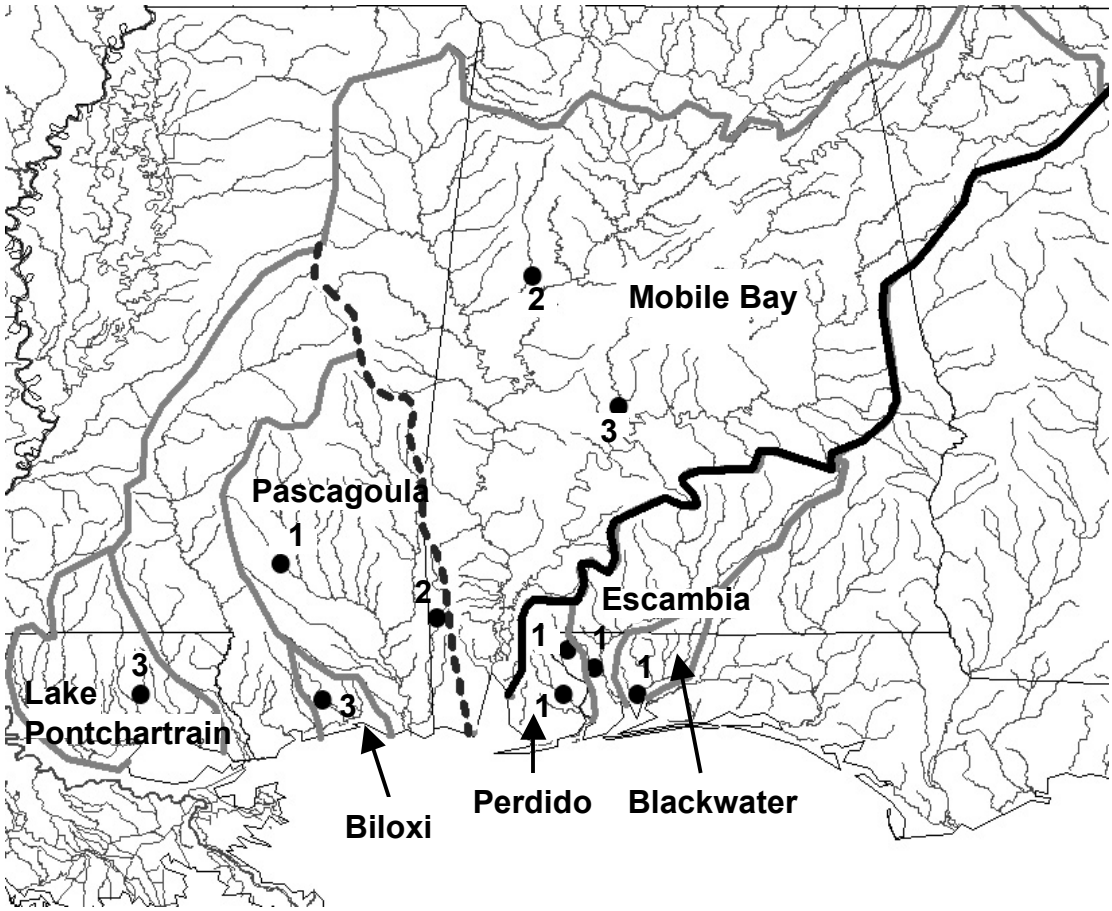


Figure 2.2. Specimens used in the phylogenetic analysis of cytochrome *b* sequence. Black dots indicate collection sites; number of specimens used from each site is indicated. Solid black line indicates boundary thought to demarcate the ranges of the two species, while dashed line indicates species boundary as indicated by mitochondrial sequence (see results).

We added an additional *A. bifascia* sequence from GenBank, and outgroup sequences of *A. meridiana, A. clara,* and *A. pellucida* (1 each), also from GenBank. Specimens for sequencing and GenBank accession numbers for sequences are given in Appendix 2.

Molecular methods – sequencing

The entire mitochondrial cytochrome *b* gene was sequenced. Nucleic acids were extracted with Qiagen DNeasy DNA extraction kits following company protocols. Extractions were then quantified on a NanoDrop ND-1000 spectrophotometer. We used the cytochrome *b* primers given in Song *et al.* (1998), and followed their PCR protocol. PCR was performed with puReTaq Ready-To-Go PCR beads (Amersham Biosciences Corp, Piscataway, NJ). Reactions contained 1.0 μL of each of the forward and reverse primers (10 pmol dilution), 2.5 units of DNA polymerase, 200 μM of each dNTP in 10 mM Tris-HCl, 50 mM KCl, and 1.5 mM $MgCl_2$ plus enough DNA sample to yield 100 ng in a total volume of 25 μL. Amplicons were purified enzymatically using ExoSap-IT (USB Corp. Cleveland, OH). We used 1/5 the manufacturer-suggested amount of enzymes for purification and incubated samples for 31 minutes at 37˚ C, followed by 15 minutes at 80˚ C to inactivate the enzymes. The gene was sequenced on an ABI 3130xl sequencer using both forward and reverse primers. The program Sequencher 4.1.1 (Gene Codes Corp.) was used to assemble the forward and reverse reads, edit ambiguous base readings, and produce consensus sequences. Alignment of the protein coding

sequences was straightforward and DNA sequences were translated into amino acids in the application Se-Al v1.0a1 to verify alignment.

Phylogenetic analysis of sequence

Both parsimony and Bayesian phylogenetic analyses were run on the cytochrome *b* sequence data. For the parsimony analysis, a heuristic search using 10,000 random stepwise-addition sequences was performed in PAUP* v4.0b10 (Swofford, 2003). Tree support was assessed from 10,000 bootstrap replicates. Bayesian analysis was run in MrBayes 3.1 (Ronquist and Huelsenbeck, 2003). MrModeltest 2.2 (Nylander, 2004) was used to determine the evolutionary model best fitting the data, with first, second, and third codon positions tested separately, and the model with the lowest AIC (Aikike Information Criterion) score chosen. Two independent runs were performed in parallel, each with four chains and temperature 0.2, for 2 million generations, with a conservative "burn in" of 2,000 generations and sampling every 100 generations.

Distance analysis of sequence

Kimura 2-parameter genetic distance was derived for every possible pairing of an *A. beanii* with an *A. bifascia* sequence in the program Phylip (Felsenstein, 2004). The average is reported as a measure of the genetic distance between the species.

Molecular methods – AFLP analysis

DNA was extracted with Qiagen DNeasy Tissue Kits and quantified on a NanoDrop ND-1000 Spectrophotometer. LI-COR AFLP protocols were followed with no modification. In the basic AFLP process, genomic DNA is digested with

*Eco*RI (a 'rare cutter' with a 6-base recognition sequence) and *Mse*1 (a 'frequent cutter' with a 4-base recognition sequence), producing millions of fragments. Because *Mse*1 cuts much more frequently than *Eco*RI, more than 90% of fragments have two *Mse*1 ends, and most of the remaining fragments have one *Eco*RI end and one *Mse*1 end. After restriction of the DNA, short known DNA sequences or "adaptors" are ligated to the fragment ends, one for ends left by *Eco*RI cutting and another for ends cut by *Mse*1. The fragments are then amplified with selective primers complementary to the cut sites and adaptors plus one additional base (A for the *Eco*RI end primer and C for the *Mse*1 end primer), in a 'preamplification' step. This step amplifies 1/16 of the fragments originally present (*i.e.*, only those that contain the specific base on each end that matches the additional base used in the primer). Fragments with two *Mse*1 ends are lost (suggested explanations given in Vos, 1995), resulting in almost exclusive amplification of fragments with one *Eco*RI end and one *Mse*1 end. Next, in the 'selective amplification,' pairs of primers are used such that one primer of each pair is complementary to the *Eco*RI cut site, adaptor, and added A plus two additional selective bases, and the other is complementary to the *Mse*1 cut site, adaptor, and added C plus two additional selective bases. Each selective primer is designated according to the type of cut (*Eco*RI or *Mse*1) to which it is complementary plus the three additional selective bases. For instance, a primer complementary to an *Mse*1 cut (and adaptor) plus the added C, plus the additional selective bases A and T would be called *Mse*1-CAT. The LI-COR kit provides eight different options for the two additional selective bases

on the *Mse*1-end primer and eight for the *Eco*RI-end primer, yielding 64 different possible primer pairs. Each primer pair produces a unique set of fragments. *Eco*RI primers are fluorescently labeled so the resulting fragments are also labeled. Samples were run on 25 cm gels (6.5% KB + acrylamide), with disposable 64-well shark tooth comb. The gel was pre-heated for 25 min at the following conditions: 1500v, 40W, 45mA and 45º C. Samples were denatured, then 0.85 µL of sample was loaded in each well and run for 3 hours 10 minutes at same conditions as pre-heat. In this study, fragments were visualized on a LI-COR Long Readir 4200 DNA sequencer, which uses a laser to detect fragments and sends a digital image to a connected computer. Two different wavelengths of dye are used, allowing sets of fragments from two different primer pairs to be visualized independently (one on each wavelength) on each gel. Primer pairs were screened using three individuals of each species until four primer pairs were found that revealed polymorphic fragments: *Eco*RI-AAG/*Mse*1-CTA, *Eco*RI-AAG/ *Mse*1-CAT, *Eco*RI-ACT/ *Mse*1-CTT, *Eco*RI-ACT/ *Mse*1-CTG. Twelve gels were scored for this project, each with bands on two wavelengths, yielding 24 gels worth of data (6 for each primer pair).

Gel scoring

"Loci" were identified by primer and length of amplified fragment, and each fragment length was considered to be a unique locus. Therefore, for example, *Eco*RI-AAG/*Mse*1-CTA 337 is assumed to designate a unique location in the genome that produces a fragment 337 base pairs long when AFLP amplification is performed with the primer pair *Eco*RI-AAG and *Mse*1-CTA. AFLP bands behave as dominant

genetic markers, where presence of a band indicates at least one copy of the relevant DNA and dominant homozygotes cannot be distinguished from heterozygotes. Absence of a specified band is interpreted as an individual being homozygous for the (recessive) genotype that does not produce the fragment in question. All interpretable loci (bands not too faint) were scored using LI-COR Saga Automated AFLP Analysis Software and rechecked manually. Presence of a band was scored as '1' and absence as '0'. Very faint or indistinct bands within loci were scored as '?'. For a given individual and primer pair, there is no reason to think that bands of consecutive length on the gel are in any way associated biologically. Therefore, it was considered extremely unlikely that an individual should be missing three or more consecutive bands present with high frequency in other individuals. Individuals lacking such consecutive bands were re-run through the entire AFLP process due to the possibility of incomplete digestion of DNA.

AFLP data analysis

The following assessments were made on the AFLP data: descriptive fragment analysis (including percent polymorphic loci, possible group-diagnostic loci, and private alleles), isolation by distance analysis, genetic diversity measured by average expected heterozygosity (Weir, 1996), genetic distance between groups, and population subdivision measured by measured as θ^B, an analog of Wright's F_{st} defined by Holsinger *et al.* (2002). In addition, an assessment was made of repeatability of the AFLP technique.

Percent polymorphic loci (and a 95% confidence interval based on 10,000 bootstrap replicates) was estimated in an Excel macro written by the author, and was evaluated at the 95% level, where a locus is considered polymorphic if the less common allele is present at a frequency of at least 5%. The Lynch and Milligan (1994) Taylor expansion method for unbiased estimation of allele frequencies from dominant marker data was used. Both possible group-diagnostic loci and private alleles were identified by eye using allele frequency output for each locus from the program Hickory (Holsinger *et al.,* 2002)**.** Possible group-diagnostic loci are loci in which one group is completely lacking an allele while the other group is fixed for that allele; private alleles are alleles present in only one group.

Isolation by distance analysis was conducted using a Mantel test in the software package TFPGA (Miller, 1997) with 999 permutations. Isolation by distance analysis compares a matrix of genetic distances between populations with a matrix of geographical distances between the localities where the populations were collected to see whether migration primarily happens between adjacent populations (genetic similarity greatest for adjacent populations). Average heterozygosity, a measure of the expected frequency of heterozygotes across all loci, and θ^B (an analog of Wright's F_{st}), a measure of the evenness of allele distribution among subgroups within a population were derived in Hickory (Holsinger *et al.,* 2002), a program specifically designed for the analysis of dominant genetic data by a Bayesian method. Credible intervals at the 95% level are reported for each of these statistics. Default values for burn in (5,000) and sampling (25,000) were used, as were all default

parameter values. In each analysis, the model assuming no within-population inbreeding (f = 0) was preferred, both because of model fit and because of lack of biological reason to suspect inbreeding. In its current implementation, Hickory cannot perform nested analyses. Nei's (1978) unbiased measure of genetic distance between groups was estimated in the software package PopGene 1.32 (Yeh *et al.,* 1997). Values reported are average pairwise distances. No way of testing for significant difference between these values is known, both because individual comparisons are nonindependent and because genetic distance is correlated with geographic distance. In all analyses of population differentiation, within drainage estimates are based on four populations within the Pascagoula drainage for western *A. beanii,* two populations within the MOB for MOB *A. beanii*, and two populations within each of the Perdido and Choctawhatchee drainages for *A. bifascia.*

Error analysis

To assess method repeatability in the current study, AFLP scores were compared between repeated runs of the same individual and rate at which scores differed between runs was reported. For the error analysis, a random selection was made of individuals that had been repeated. Because many of these individuals had been rerun because of concerns about quality, individuals with 10% or more loci scored as "uncertain" for two or more of the four primer pairs were excluded as untrustworthy. An Excel macro written by Michael Tourtellot (University of Kansas) was used to compare repeated runs.

Results

Phylogenetic analysis

The 1140 bases of the cytochrome *b* gene from 23 individuals of *Ammocrypta* contained 240 variable sites (24 in first positions, 2 in second positions, and 214 in third positions) of which 122 sites were phylogenetically informative (11 in first, 0 in second, and 111 in third positions respectively). There was evidence for base composition bias across all three codon positions (0.219 A, 0.287 T, 0.326 C, 0.169 G), which was more pronounced in third positions (0.234 A, 0.228 T, 0.446 C, 0.091 G). The average transition:transversion ratio was 3.42. Plots of Tamura-Nei distance versus total transitions and transversions for all positions showed a linear relationship, suggesting no evidence of site saturation (data not shown).

In both the parsimony and Bayesian analyses, MOB specimens of *A. beanii* form a monophyletic clade that grouped with a clade containing all *A. bifascia* individuals. Figure 2.3 shows the 50% majority rule tree from the Bayesian analysis, which is identical in ingroup topology to the parsimony tree. A separate clade is recovered containing all other *A. beanii* individuals. In the parsimony analysis, bootstrap support for each of these clades is strong, with 86% support for the clade of MOB *A. beanii,* 92% for the clade of *A. bifascia* + MOB *A. beanii,* and 92% for the clade of *A. beanii* excluding the MOB *A. beanii*. In the Bayesian analysis, SYM + I was run for first positions (I = 0.8466, equal rates for all sites), HKY was run for second positions (equal rates for all sites, proportion of invariable sites = 0), and GTR + G was run for third positions (gamma distribution shape parameter = 1.2703,

proportion of invariable sites = 0). The Bayesian analysis agrees exactly with the parsimony analysis in topology. Credibility values were higher than bootstrap values

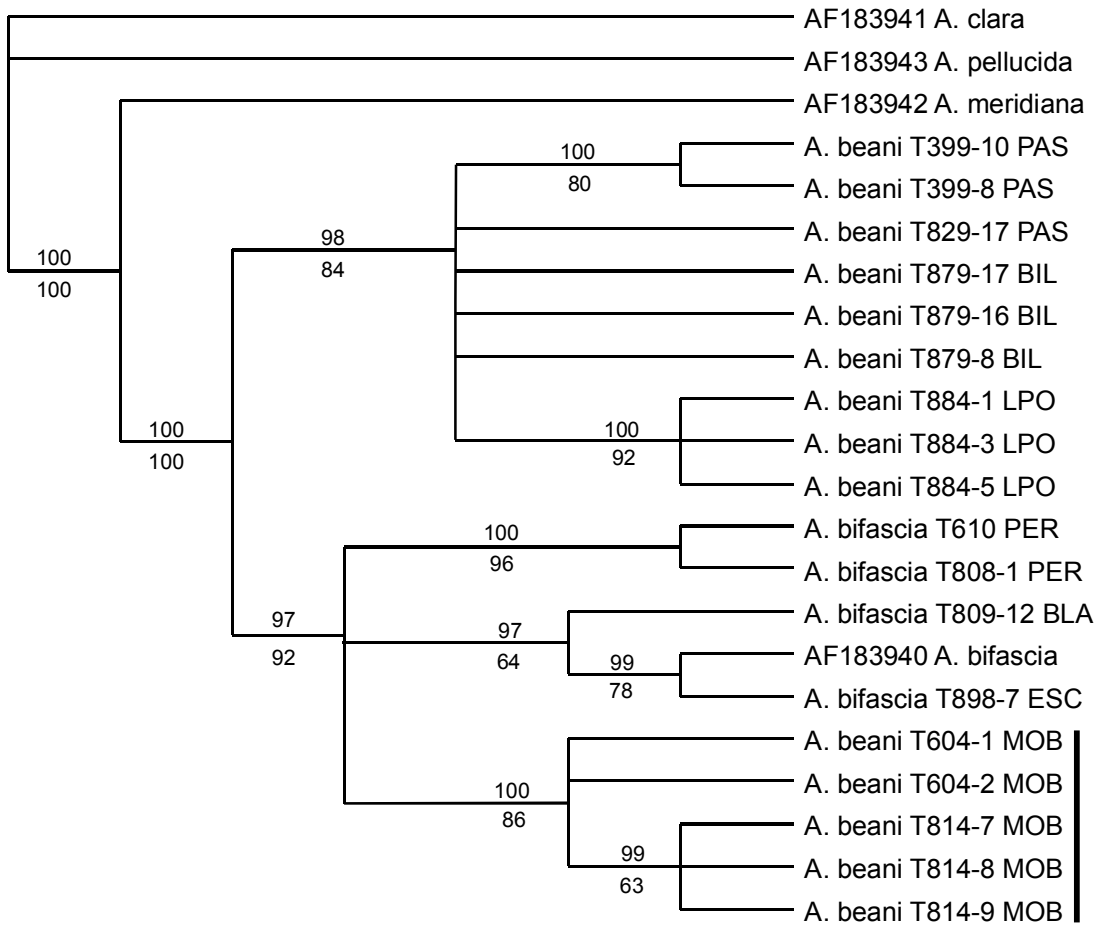


Figure 2.3. Phylogenetic tree based on cytochrome *b* mitochondrial sequence. Parsimony bootstrap values (10,000 replicates) are placed below nodes, while Bayesian credibility values are placed above nodes. The sequence data was analyzed by parsimony in PAUP* (Swofford, 2003) and also by Baysian analysis in the program MrBayes (Ronquist and Huelsenbeck, 2003). Topologies were in exact concordance in both analyses. Nodes of less than 95% probability (Bayesian) or less than 50% bootstrap recovery (parsimony) were collapsed. Drainages from which individuals were used were Lake Pontchartrain (LPO), Biloxi (BIL), Pascagoula (PAS), Mobile Bay (MOB), Perdido (PER), Escambia (ESC), and Blackwater (BLA). Drainage abbreviations are as used in Swift *et al.* (1986). Black bar highlights location of MOB specimens.

in the parsimony analysis. Support was 100%, 97%, and 98% for MOB *A. beanii,* [MOB *A. beanii* + *A. bifascia],* and all non-MOB *A. beanii,* respectively.

Partitions of the data

Because the phylogenetic affinity of the MOB *A. beanii* was uncertain, population genetic analyses were run on the data partitioned in several ways: (A) along accepted species boundaries (ranges as in figure 2.4A), (B) divided into 3 groups with MOB *A. beanii* separate from both western *A. beanii* and *A. bifascia* (ranges as in figure 2.4B), and (C) with MOB *A. beanii* and *A. bifascia* treated as a clade together, sister to western *A. beanii* (ranges as in figure 2.4C).

Distance analysis of sequence

Average pairwise Kimura 2-parameter distance between cytochrome *b* sequences of *A. beanii* individuals and *A. bifascia* individuals was 0.0117. When the specimens were partitioned according to the results of the phylogenetic analysis ([MOB *A. beanii* + *A. bifascia]* vs western *A. beanii),* average pairwise distance was 0.0144. Distance between western *A. beanii* and MOB *A. beanii* was 0.0147.

Fragment analysis – polymorphic loci, diagnostic loci, private alleles

A total of 223 bands were scored, of which 17 were completely monomorphic and 17 others had rare allele frequency less than 5%, making 84.8% of loci polymorphic across the entire data set using the 95% criterion. Polymorphism levels for subgroups within the data are shown in figure 2.5 (lower numbers) for different partitionings of the data. As mentioned above, analyses are made for three alternative

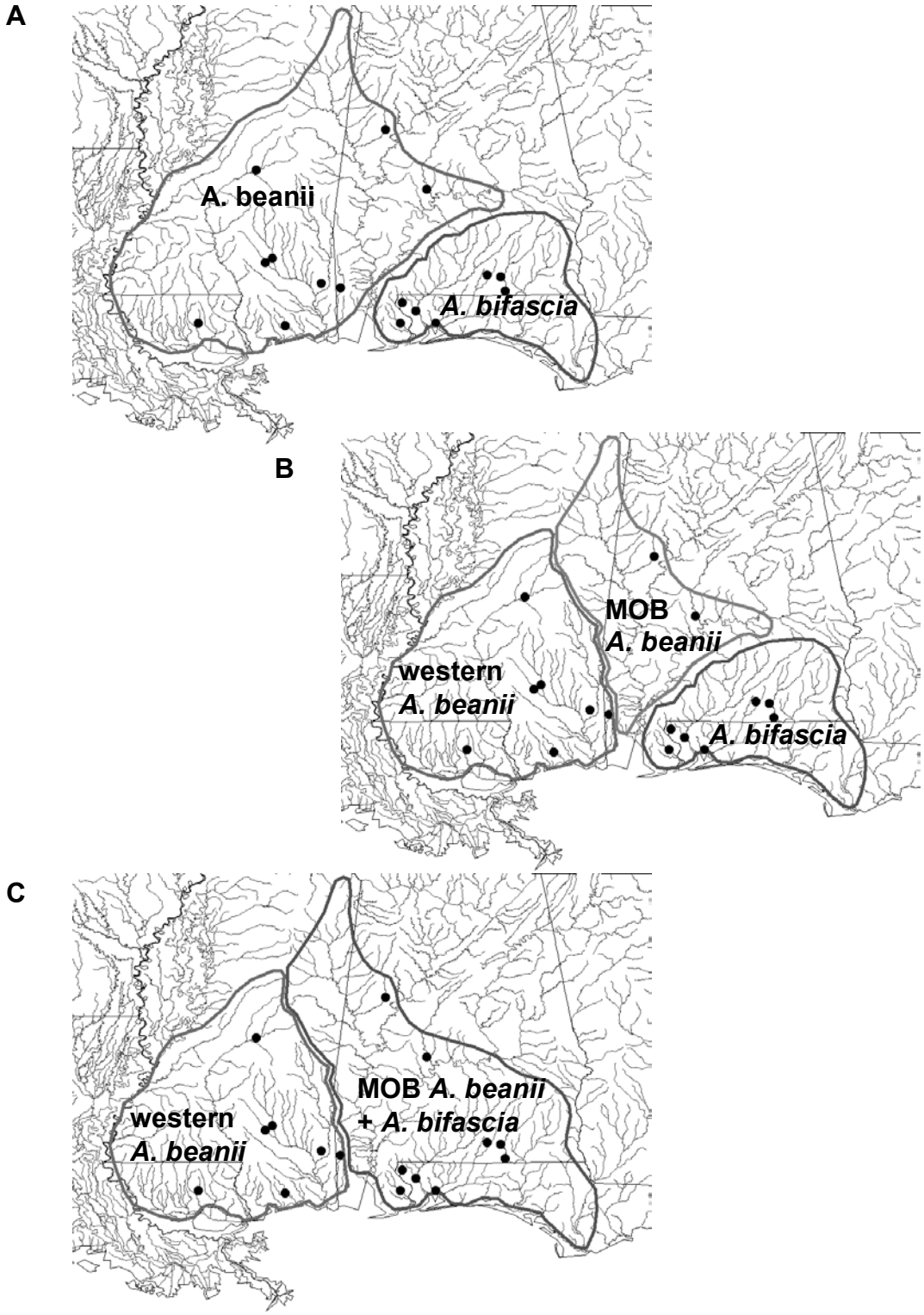


Figure 2.4. Group ranges in three different partitions. (A) accepted species boundaries; (B) MOB a separate group; (C) MOB grouped with *A. bifascia.*

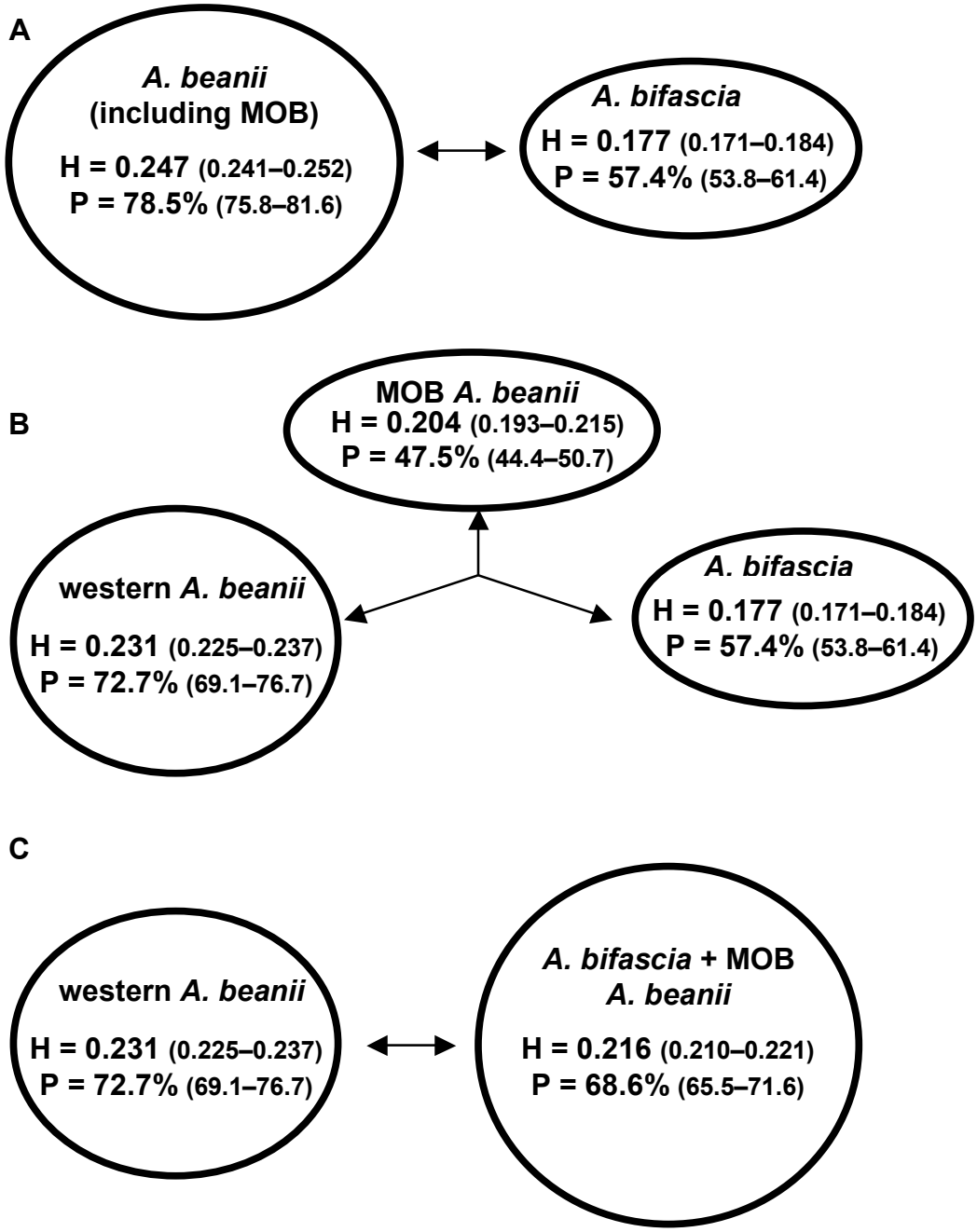


Figure 2.5. Heterozygosity and polymorphism for different partitions and subpartitions of the data. Upper number is heterozygosity (and 95% credible interval) and lower number is percent polymorphic loci at the 95% level (most frequent allele frequency < 95%) with 95% confidence interval shown. Heterozygosity was calculated in the program Hickory (Holsinger *et al.,* 2002) and polymorphism in an Excel macro written by the author.

partitionings of the data. When the individuals are divided into accepted species groupings, *A. beanii* has 78.5% polymorphic loci and *A. bifascia* has 57.4% polymorphic loci (figure 2.5A). When MOB *A. beanii* is separated from western *A. beanii,* it has the lowest percent polymorphic loci at 47.5%, while western *A. beanii* is higher at 72.7% (figure 5B). If MOB *A. beanii* is placed in a group with *A. bifascia* (the grouping supported by the phylogenetic analysis), this group has 68.6% polymorphic loci (figure 2.5C).

There were no bands diagnostic for "species" (*i.e.*, for any of the five groups *A. beanii,* western *A. beanii,* MOB *A. beanii, A. bifascia,* and *[A. bifascia* + MOB *A. beanii]),* although four bands came close to being diagnostic (table 2.1); three of which would diagnose groups in the partition *[A. beanii* (including MOB)] vs *[A. bifascia]* and one of which (table 2.1, row 2) would diagnose groups in the partition [western *A. beanii]* vs [MOB *A. beanii* + *A. bifascia].* In addition, 25 bands showed similar frequency (within 5%) for two among the three groups (western *A. beanii,* MOB *A. beanii,* and *A. bifascia)* and differed in the third by more than 25%. Eleven bands had similar frequency in western *A. beanii* and MOB *A. beanii,* ten for MOB *A. beanii* and *A. bifascia,* and four for western *A. beanii* and *A. bifascia.* There were a total of six bands for which one group had a private allele: three for western *A. beanii* and three for *A. bifascia.*

Isolation by distance analysis

Results of the isolation by distance analysis are shown in figure 2.6. As shown by the Mantel test, correlation is fairly strong and significant for *A. beanii*

Table 2.1. Fragments nearly diagnostic for groups. Shown are proportions of individuals for which the band is present. Notice fragment in row two diagnoses the groups [western *A. beanii]* vs [MOB *A. beanii* + *A. bifascia],* while the other three fragments diagnose *[A. beanii* (including MOB)] vs *[A. bifascia].*

#	Fragment	western *A. beanii*	MOB *A. beanii*	*A. bifascia*
1	*Eco*RI-AAG/*Mse*1-CTA 337	97/99	19/20	0/113
2	*Eco*RI-AAG/*Mse*1-CAT 141	99/101	0/20	1/115
3	*Eco*RI-AAG/*Mse*1-CAT 143	3/95	0/20	115/115
4	*Eco*RI-AAG/*Mse*1-CAT 211	2/100	1/18	113/113

when MOB is included (r =0.81, p =0.001, figure 2.6A) but much weaker and just bordering on significant for western *A. beanii* alone (r = 0.51, p = 0.06, figure 2.6B). Correlation is significant for *A. bifascia* (p = 0.02, figure 2.6C), although the correlation coefficient value is only 0.34.

Population genetic analysis – genetic diversity

Figure 2.5 shows genetic diversity measured as average heterozygosity (H_T) for the three different possible partitionings of the data. Figure 2.5B shows that western *A. beani* has the highest heterozygosity (0.231) when specimens are partitioned into three groups, followed by MOB *A. beanii* at 0.204 and *A. bifascia* at 0.177. Unlike in the case of percent polymorphic loci, the value is not lowest for MOB *A. beanii.* When the group MOB *A. beanii* is added to either western *A. beanii* (making the group *A. beanii* – figure 2.5A vs 2.5B) or *A. bifascia* (making the group [MOB *A. beanii* + *A. bifascia]* – figure 2.5C vs 2.5B) heterozygosity increases over

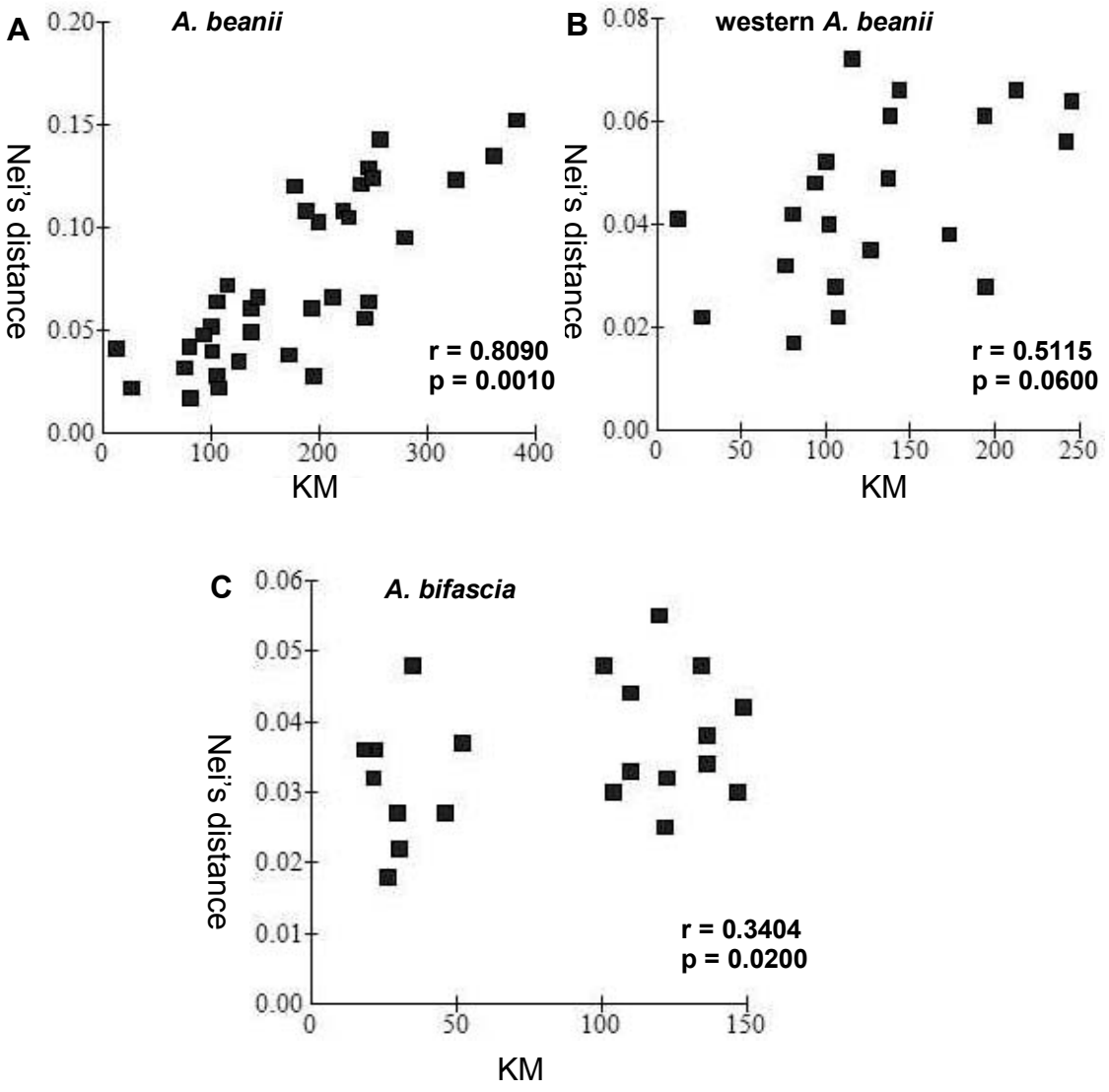


Figure 2.6. Plot of Nei's genetic distance (pairwise comparisons of all individuals between populations) versus geographical distance between populations as assessed in a Mantel test (A) for *Ammocrypta beanii* including MOB; (B) for western *A. beanii;* and (C) for *A. bifascia.*

the level it was in that group by itself (to 0.247 for *A. beanii* and to 0.216 for [MOB *A. beanii* + *A. bifascia]).* Table 2.2 shows heterozygosities found at the drainage and population level. Average heterozygosity at the drainage level and population level are 0.209 (range 0.207–0.211) and 0.204 (range 0.187–0.211), respectively, for *A.*

Table 2.2. Heterozygosity at the population and drainage level. Many drainages are only represented by one population, so there is no estimate of drainage-level heterozygosity. 95% credible interval given in parentheses. Heterozygosities were calculated in the program Hickory (Holsinger *et al.*, 2002).

Drainage	Heterozygosity	Population	Heterozygosity
LPO		LPO	0.187 (0.173–0.201)
PRL		PRL	0.223 (0.211–0.234)
BIL		BIL	0.223 (0.212–0.233)
PAS	0.207 (0.199–0.215)	PAS1	0.197 (0.185–0.209)
		PAS2	0.187 (0.175–0.200)
		PAS3	0.196 (0.184–0.209)
		PAS4	0.216 (0.204–0.227)
MOB	0.211 (0.201–0.222)	MOB1	0.199 (0.187–0.210)
		MOB2	0.211 (0.197–0.226)
A. beanii average	0.209	*A. beanii* average	0.204
PER	0.151 (0.141–0.160)	PER1	0.147 (0.137–0.157)
		PER2	0.155 (0.142–0.166)
ESC		ESC	0.152 (0.143–0.160)
BLA		BLA	0.145 (0.133–0.166)
YEL		YEL	0.167 (0.158–0.177)
CHO	0.150 (0.141–0.159)	CHO1	0.150 (0.139–0.160)
		CHO2	0.147 (0.135–0.158)
A. bifascia average	0.150	*A. bifascia* average	0.152

beanii and 0.150 (range 0.150 to 0.151) and 0.152 (range 0.147 to 0.167), respectively, for *A. bifascia.*

Population genetic analysis – genetic distance

Figure 2.7 shows Nei's (1978) average unbiased pairwise genetic distance between populations from different groups, as well as the average and range of pairwise distances between drainages within each group, and average distance between populations within drainages (in parentheses) in the different data partitions.

Partition A shows an average distance of 0.163 between populations of the accepted *A. beanii* and *A. bifascia* groups with larger inter- and intra-drainage distances in *A. beanii* than in *A bifascia* (0.085 vs 0.027 and 0.043 vs 0.026, respectively). When individuals in each species were clustered into one group (not treated as separate populations) Nei distance between the two species was 0.114. This number is presented for later comparison with sequence divergence, which is derived without division of individuals into populations.

In partition B, average pairwise distance of MOB *A. beanii* populations from western *A. beanii* populations is 0.117, with a range of 0.094–0.150, while pairwise distances between MOB *A. beanii* populations and *A. bifascia* populations are larger, averaging 0.167 with a range of 0.132–0.201. In this partition, average distance between drainages is larger for western *A. beanii* than for *A. bifascia* (0.047 vs 0.027), while distances between populations within drainages are very similar for western *A. beanii* and *A. bifascia* (0.025 vs 0.026). The distance between the two populations sampled for MOB is relatively high (0.061), even higher than the distance between drainages in the other two groups. In partition C, the [MOB *A. beanii* + *A. bifascia]* group has the highest average between-drainage pairwise distance of any group in any partition (0.095 with a range of 0.016–0.201). There is an average pairwise distance of 0.149 between populations in this group and western *A. beanii* populations with a range of 0.094–0.194). When individuals in each of these two groups were clustered into one group (not treated as separate populations) Nei's distance between the two groups was 0.092. Again, this number is presented for later

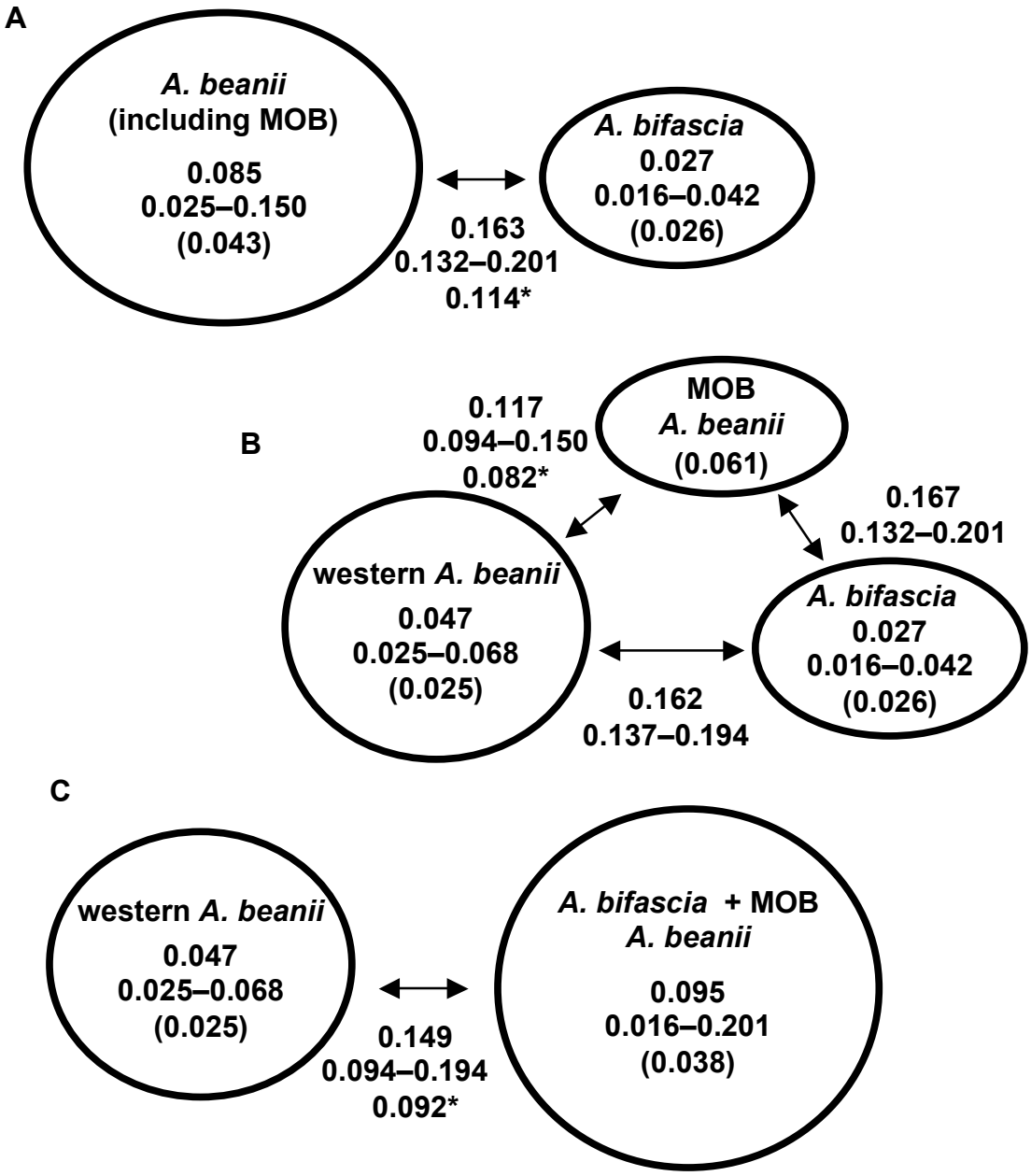


Figure 2.7. Nei's (1978) unbiased genetic distances between groups calculated in PopGene 1.32 (Yeh *et al.,* 1997). Numbers in figure are average pairwise distances between drainages either between groups (by arrows) or within them (inside circles). Minimum and maximum between-population distances are also given. Numbers in parentheses are average pairwise distance between populations within drainages for the three groups. Starred values are distances between groups when individuals within groups are not divided by drainage. A, B, and C represent different partitionings of individuals (discussed in text).

comparison with sequence divergence.

Population genetic analysis – θ^B

Figure 2.8 shows genetic differentiation (measured as θ^B, an analog of Wright's F_{st}) at three different levels of analysis: the most inclusive level of group (species or other large group) within total; the less inclusive level of drainage within group; and the least inclusive level, population within drainage. At the most inclusive level, genetic subdivision is greatest when specimens are divided into three groups (0.277 – figure 2.8B), it is least in the partition supported by cytochrome *b* phylogeny (0.209 – figure 2.8C), and is intermediate when specimens are partitioned into accepted species (0.261 – figure 2.8A). These differences are only significant when comparing partition B (the largest θ^B) with partition C (the smallest θ^B). Subdivision between drainages is highest (0.243) in partition C for the group [MOB *A. beanii* + *A. bifascia],* next highest (0.153) for *A. beanii* (including MOB – partition A), and lowest in the groups western *A. beanii* and *A. bifascia* (0.075 and 0.174, respectively). The highest value for θ^B at the level of population within drainage (0.099) is in the MOB group, while the lowest values are for western *A. beanii* and *A. bifascia* (0.022 and 0.041, respectively). For all partitions θ^B at the level of population within drainage is less than that at the level of drainage within larger group, with the exception in partition B that θ^B between populations in the MOB drainage is higher than that between drainages in the other groups.

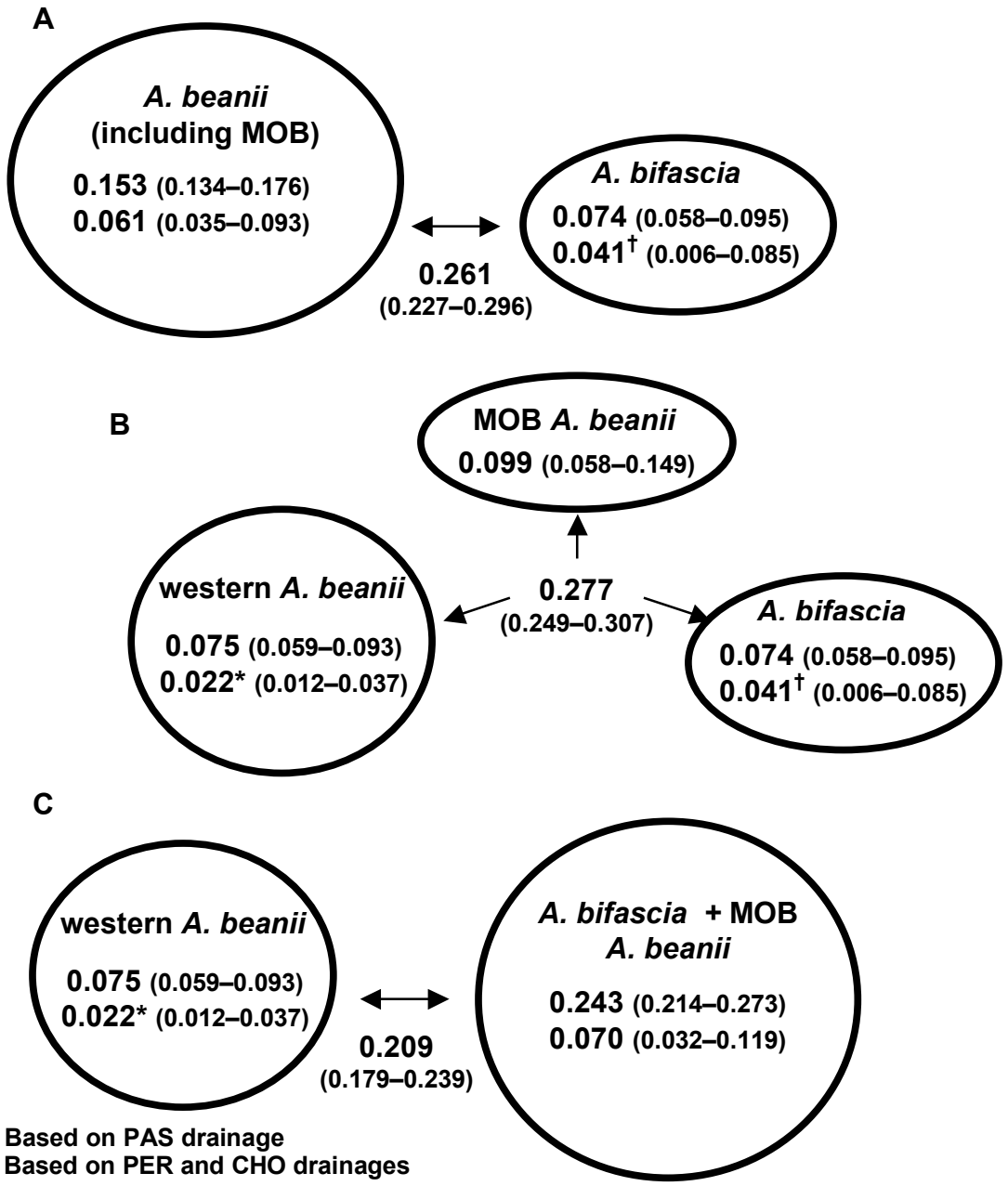


Figure 2.8. θ^B for different partitions and subpartitions of the data calculated in the program Hickory (Holsinger *et al.,* 2002). θ^B between larger groups shown next to arrows between groups (represented by circles). Within groups, upper number is drainage within larger group and lower number is population within drainage, except in part B for MOB *A. beanii,* where the one number given is for population within drainage. Numbers in parentheses are 95% credible intervals. A, B, and C represent different partitionings of individuals (discussed in text).

Error analysis

A total of 55 individuals were rerun through the AFLP process, of which 38 were randomly selected for the repeatability analysis. Twenty of these were excluded by the 10% question mark cutoff. Based on the 18 remaining individuals, 4.6% of scores (out of a total of 3153 scores compared) differed between runs.

Discussion

Phylogenetic analysis

The phylogenetic analysis (figure 2.3) indicates that the *Ammocrypta* specimens in this study fall in two main groups, one containing western *A. beanii* and the other containing MOB *A. beanii* and *A. bifascia.* If the specimens in this analysis are considered to form two species, this analysis indicates that the dividing line should be between the western *A. beanii* and the group [MOB *A. beanii* + *A. bifascia]* (figure 2.4C).

Distance analysis of sequence

Level of cytochrome *b* sequence divergence indicates a much more recent separation of groups than the posited vicariant event of 2–4 million years ago. There is a divergence of 0.0117 in cytochrome *b* sequence between *A. beanii* and *A. bifascia* as the species are currently recognized. This implies that each species is 0.0059 dissimilar from the common ancestor. The geographic distribution of these groups is such that they could have formed during the vicariant event hypothesized to have taken place across the eastern boundary of the MOB, however this event is thought to have taken place 2–4 million years ago. Even if the event took place only 2 million

years ago, this amount of sequence divergence translates to a rate of sequence evolution of 0.295% per million years, which is a very low number given what is known about the rate of cytochrome *b* evolution in other taxa. It is generally thought that the cytochrome *b* gene evolves at a rate of 1 to 2 percent per million years in poikilotherms (Thorp *et al.,* 2005), although the exact rate varies by taxonomic group, seemingly with some correlation to body size, body temperature, and even clutch size (Bromham, 2002; Gillooly *et al.,* 2005). Poulkakis *et al.* (2005) estimated the rate of cytochrome *b* evolution to be 1.55% per million years in lizards of the genus *Podarcis,* while Guiking *et al.* (2006) found a general mitochondrial evolution rate of 1.35% per million years for *Natrix* snakes. Domingues *et al.* (2005) found a higher rate of cytochrome *b* evolution at 2.5% per million years in *Chromis* fishes, however. Given that different taxa exhibit different rates of sequence evolution, it is not possible to use sequence divergence to pinpoint a time of speciation without some way to calibrate rate of evolution for the taxon in question. However, based on the studies above, it is probable that *Ammocrypta* cytochrome *b* evolves at a rate falling within the range of 1.5% to 2.5% per million years. Such a range would place the speciation event within the last 400,000 years. If, on the other hand, the division lies between [MOB *A. beanii* + *A. bifascia]* and western *A. beanii,* as indicated by the cytochrome *b* phylogeny, the divergence between groups is 0.0144, giving each group a divergence of 0.007 from the common ancestor. Assuming a rate of evolution of 1.5% per million years or greater would place this divergence at 470,000 years or

less. Both estimates are considerably more recent than the age hypothesized for the vicariant event.

Fragment analysis

'Fragment analysis' included descriptive measures on the AFLP fragments such as percent polymorphic loci and identification of group-diagnostic loci. Analysis of levels of polymorphism (proportion of loci at which the frequency of less common allele is at least 5%) in different groupings of individuals supports a closer relationship between MOB *A. beanii* and western *A. beanii* than between MOB *A. beanii* and *A. bifascia.* The group western *A. beanii* has nearly as high a level of polymorphism as the group *A. beanii* (western + MOB *A. beanii)* (figure 2.5B compared to 2.5A), indicating that the two groups western and MOB *A. beanii* commonly agree (though not always) on presence or absence of bands at loci. The combined group has slightly higher polymorphism than either constituent group due to the presence of some opposite loci (a band frequently present in one group and only rarely or not at all in the other). That the polymorphism of the group [MOB *A. beanii* + *A. bifascia]* is at least 10% higher than either MOB *A. beanii* or *A. bifascia* considered separately (figure 2.5C vs 2.5B) indicates that these groups differ in presence versus absence of bands for more loci than do the groups MOB *A. beanii* and western *A. beanii.* The low percent polymorphic loci estimate for MOB *A. beanii,* when compared to the groups western *A. beanii* and *A. bifascia* could well be an artifact of sample size, since the MOB *A. beanii* are only represented by two populations. Lower rate of polymorphism in *A. bifascia* compared to western *A.*

beanii is likely to be biologically meaningful, however, since comparable numbers of drainages, populations, and individuals were sampled for each of these groups (4, 7, and 101 vs 5, 7, and 111, respectively).

Because of the error rate in the AFLP analysis, it is unlikely that an absolutely group-diagnostic locus could be detected by this method of analysis. While no absolutely diagnostic loci were found, however, the four nearly group-diagnostic bands listed in table 2.1 allow a high probability of assigning unknown individuals to the correct group based on genotype no matter which of the possible data partitions is considered. Based on the three bands relevant to these groups (bands 1, 3 and 4 in table 2.1), there is a 99.9% chance of correct assignment between *A. beanii* and *A. bifascia*, and also a 99.9% chance between MOB *A. beanii* and *A. bifascia* (table 2.3). This is based on the proportion of individuals for which group identity is correctly indicated by each fragment and assuming an individual will be assigned to a group whenever the majority of bands so indicate. Group identity is considered to be

Table 2.3. Probability of correct assignment of individuals between different groups based on nearly diagnostic loci listed in table 2.1.

Groups to distinguish between	Assignment probability
A. beanii (including MOB) and *A. bifascia*	99.9%
MOB *A. beanii* and *A. bifascia*	99.9%
Western *A. beanii* and MOB *A. beanii*	99.0%
Western *A. beanii* and [MOB *A. beanii* + *A. bifascia]*	98.7%

'correctly indicated' when an individual possesses the character state most common for the group to which it belongs. Since three fragments are nearly diagnostic between *A. beanii* and *A. bifascia*, the probability of correct assignment between these groups, for example, is equal to the proportion of all individuals for which all three fragments are expected to correctly indicate group plus the proportion of all individuals for which two fragments are expected to correctly indicate group. Only fragment 2 in table 2.1 can distinguish western *A. beanii* individuals from either MOB *A. beanii* individuals or [MOB *A. beanii* + *A. bifascia*] individuals. Based on this one locus, individuals can be correctly assigned between western *A. beanii* and MOB *A. beanii* 99.0% of the time, and between western *A. beanii* and [MOB *A. beanii* + *A. bifascia]* 98.7% of the time. These results are useful because Wiley and Hagen (1997) were unable to distinguish genetically between *A. beanii* and *A. bifascia* specimens based on cytochrome *b* sequence.

Isolation by distance analysis

The isolation by distance analysis indicates that MOB *A. beanii* and western *A. beanii* groups are quite distinct. As shown in figure 2.6, when isolation by distance analysis is performed on these two groups combined, populations in the analysis are distinctly isolated by distance. If the MOB *A. beanii* populations are removed from the analysis, the isolation by distance greatly decreases, implying that western *A. beanii* populations have enough migration between them to maintain fairly homogenous allele frequencies or that *A. beanii* populations have colonized the area

fairly recently from a common source, but there is not that level of relatedness between the western and the MOB populations.

Population genetic analysis – genetic diversity

Analysis of group heterozygosities supports an affinity between MOB *A. beanii* and western *A. beanii.* Heterozygosity estimates increase as allele frequencies get closer to 0.5. This happens when MOB *A. beanii* is added to either western *A. beanii* or *A. bifascia* (figure 2.5A and 2.5C versus 2.5B) indicating that allele frequencies in these groups often differ from those in MOB such that combining the groups pulls the allele frequencies at many loci closer to 0.5 than it was in either group, and meaning they originally differ from 0.5 in opposite directions, meaning they are genetically distinct in terms of allele frequencies. The greater increase when MOB *A. beanii* is added to *A. bifascia* than to western *A. beanii* indicates that MOB *A. beanii* differs more in allele frequencies across loci from *A. bifascia* than from *A. beanii*. Heterozygosities at the population level vary more than those at the drainage level, which may be a result of the generally smaller size of the population samples.

Population genetic analysis – genetic distance

The main pattern that emerges from the genetic distance analysis is that the MOB *A. beanii* populations are distant from all other populations in the analysis. Though sampling was not as extensive in MOB as in the other two groups, the relatively large distance between MOB *A. beanii* populations and other populations in the study is likely real and not just an artifact of small sample size. One of the MOB *A. beanii* populations had a respectable sample size of 15 (figure 2.1), and the

distance between this population and any of the populations in western *A. beanii* was at least twice as large as the largest pairwise distance between drainages within western *A. beanii.* This distance is reflected in figure 2.7B, which shows that the smallest pairwise distance between western *A. beanii* populations and either of the MOB populations (0.094) is still considerably larger than the largest pairwise distance between drainages within western *A. beanii*. The distance analysis also supports a closer relationship between MOB *A. beanii* and western *A. beanii* (distance 0.117) than between MOB *A. beanii* and *A. bifascia* (distance 0.167). The relatively large distance between populations within the MOB may well be an artifact of small sample size, however, since the other MOB population only had six sampled individuals.

The distance analysis best supports dividing individuals according to partition B. Distances between the three main groups in partition B are large compared to distances between drainages (or populations) within groups (figure 2.7B), indicating that this partition is placing similar genotypes together and separating dissimilar ones. Within-group distances increase in the other partitions, indicating that the other partitions do not group genotype similarity and partition genotype difference as effectively. Partition C (which is best supported by the phylogenetic analysis) has the highest within-group distances and the lowest between-group distance and thus does the worst job of grouping genotypes.

Comparison can be made between the estimate of divergence from the nuclear AFLP data and the estimate of divergence derived from the mitochondrial

cytochrome *b* sequence. Nei's distance, calculated for the AFLP data, is approximately $2\mu t$ where μ is the infinite alleles mutation rate at the loci examined. Based on sequence divergence, time of divergence between *A. beanii* and *A. bifascia* was estimated to be 400,000 years or less. Given this divergence time and the between-species Nei distance calculated earlier (0.114) the mutation rate would be 1.4×10^{-7} mutations per year per locus for these AFLP loci. Assuming that these mutations are allele loss, and given that any one of ten DNA sites may change to prevent a fragment from being recovered (6 in the *Eco* RI recognition site and 4 in the Mse 1 recognition site), this translates to a mutation rate of 1.4×10^{-8} changes per nucleotide site per year (or generation, since the generation time of these fishes is one year). This is somewhat larger than the genomic mutation rate estimates given in Drake *et al.* (1998), even when these rates are corrected to represent mutation rate per generation. This correction is done by multiplying their estimated mutation rate per cell division by the estimated number of divisions involved in gamete production. Their numbers, once this correction is made, are 8.5×10^{-9} mutations per base per generation for *Drosophila melanogaster* and 2.0×10^{-9} mutations per base per generation for *Caenorhabditis elegans*. Using the rate of 8.5×10^{-9} (the higher rate) as a reasonable estimate for *Ammocrypta* genomic mutation rate yields a divergence time of 670,000 years, which is 68% larger than the timeframe indicated by the mitochondrial sequence analysis. Using the lower number would yield an older estimated divergence time, and an even bigger discrepancy. A more distant divergence time estimate from nuclear data than from mitochondrial data, such as

seen here, could be explained by if there was more introgression of mitochondrial genes than nuclear during or post speciation (Coyne and Orr, 2004), although the error in these mutation rate estimates may be large enough that this discrepancy is only an artifact.

Population genetic analysis – θ^B

The highest θ^B at the most inclusive level is found when specimens are divided into partition B (western *A. beanii,* MOB *A. beanii,* and *A. bifascia* – figure 2.8). The increase over partition A is not significant while the increase over partition C is. This increase suggests that this partition best divides groups of distinct allele frequencies. In addition θ^B between drainages drops significantly from 0.153 in *A. beanii* when MOB drainage is included to 0.075 among western *A. beanii* drainages only. This indicates that there is less genetic differentiation among the western *A. beanii* drainages than among those drainages and the MOB *A. beanii*, and that western *A. beanii* form a distinct group to the exclusion of the MOB specimens. θ^B is also significantly lower in *A. bifascia* than in the group *[A. bifascia* + MOB *A. beanii]*. These comparisons show that the division into three groups may best reflect patterns of genetic differentiation in the AFLP data.

Error analysis

The minority of studies using AFLP data present an analysis of error rate of the method, but a sample of 13 that did present error data revealed that anywhere from 0 to 7.9% of scores differ between repeated runs of the same individual (Tohme *et al.,* 1996; Hartl and Seefelder, 1997; Hongtrakul *et al.,* 1997; Donaldson *et al.,*

1998; Busch *et al.,* 2000; Donaldson *et al.,* 2000; Bagley *et al.,* 2001; Congiu *et al.,* 2001; Goulao *et al.,* 2001; Miller *et al.,* 2002; Mock *et al.,* 2002; Kruse *et al.,* 2003; Whitehead *et al.,* 2003). Some studies also found that DNA of some individuals fails to work in the AFLP process (Palacios *et al.,* 1999; Busch *et al.,* 2000; Donaldson *et al.,* 2000; Bensch and Åkesson, 2005). The 4.6% level of non-repeatability found in this study probably had a small effect on the analyses performed on the AFLP data. If there is an equal probability that any given score is incorrect the effect would be to bias all allele frequencies in the direction of 50% (given actual genotypes at a locus, it is more likely that one of the more common score types will be observed incorrectly), increasing polymorphism and decreasing θ^B and distances between populations and groups. Simulation studies could be performed to assess the effect such an error rate is likely to have.

Comparison of sequencing and AFLP results

The phylogenic analysis based on cytochrome *b* sequence and the AFLP work indicate very different affinities for the MOB *A. beanii.* The phylogenetic analysis indicates that MOB *A. beanii* group with the *A. bifascia* and are not related to the western *A. beanii.* The AFLP work, on the other hand, indicates that western *A. beanii,* MOB *A. beanii,* and *A. bifascia* form three distinct groups, with the MOB *A. beanii* closer to the western *A. beanii* then to the *A. bifascia.* There are at least two possible explanations for this disparity. The phylogenetic analysis could be giving a false picture of these groups because of lineage sorting issues, or the AFLP analysis,

which is based on distance, may simply not be an accurate assessment of group relationships.

In the first of these possibilities, the mitochondrial cytochrome *b* gene could be a false representation of phylogeny due to the vagaries of lineage sorting or to mtDNA capture. In lineage sorting, if the common ancestor of all three groups had two distinct mitochondrial types that were passed down to all three groups, it is possible by random chance that the MOB *A. beanii* and the *A. bifascia* happened to retain descendants of one haplotype, while western *A. beanii* retained the other (figure 2.9). This occurrence would cause MOB *A. beanii* and *A. bifascia* to cluster in a phylogeny based on mitochondrial sequence, even if MOB *A. beanii* had actually diverged more recently from western *A. beanii.* Nei (1987) derived a formula for the probability (p) that lineage sorting would give a gene tree different from that of the species or population tree. The formula is $p = 2/3(e^{-T/2N})$. $T = t_1 - t_2$, where t_1 is time of divergence of the more distant group and t_2 is time of the splitting of the two sister groups (see figure 2.9), and N is the effective population size. This probability is high when time between divergences (T) is small and effective population size is large. An estimate of t_1 at 670,000 years was calculated from AFLP data in the 'genetic distance' part of the Discussion Section. Using the same method (assuming a genomic mutation rate of 8.5×10^{-9} base changes per generation), the AFLP Nei's distance of 0.082 between western and MOB groups of *A. beanii* yields a t_2 estimate of 480,000 years. This yields T of 190,000 years. Using this estimate for T, one can calculate how large N would have to be to correspond to various probability levels

that the gene tree differs from the population tree. Table 2.4 gives values of N corresponding to several values of p. To have a probability of 0.05 or greater, N must be greater than 37,000. It is difficult to say whether N of these fish may have been this high or higher because very few estimates of effective population size have been

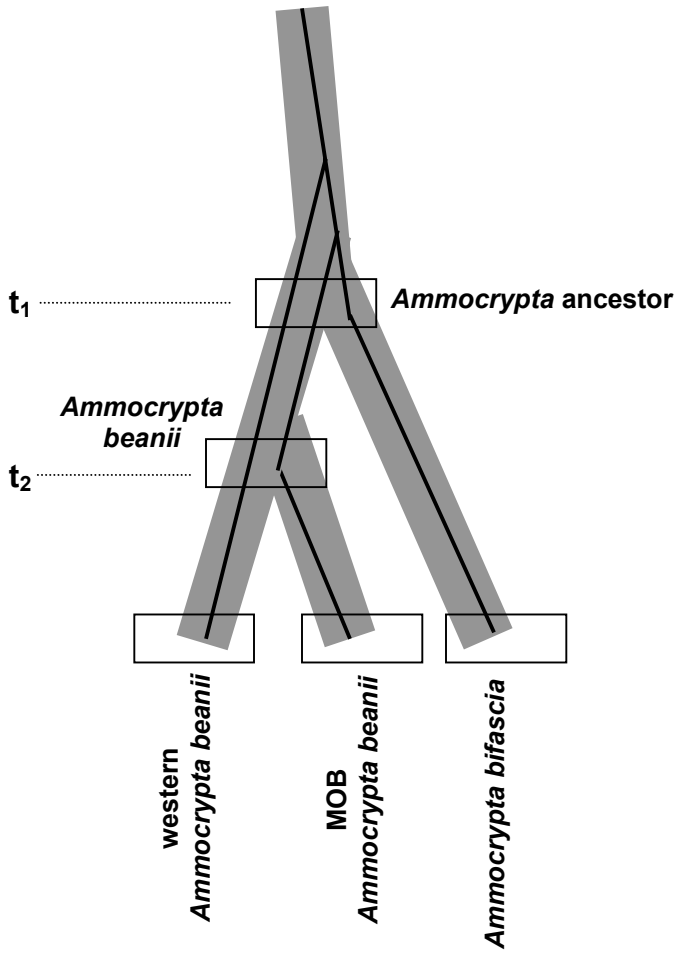


Figure 2.9. Lineage sorting. Grey lines represent evolutionary lineages of populations; black lines represent evolutionary lineages of genes. Notice the gene lineage in MOB *A. beanii* is most closely related to the gene lineage in *A. bifascia,* while at the population level the group MOB *A. beanii* is most closely related to western *A. beanii.* The times t_1 and t_2 represent the times of divergence of *A. beanii* from *A. bifascia* and of MOB from western *A. beanii*, respectively, at the population level.

Table 2.4. Effective population size (N) estimates corresponding to different probabilities (p) that gene tree will differ from population tree. These population sizes are derived from Nei's (1987) equation relating N, p, and time between population-level divergence events (see text). Time between divergence events for the groups treated here is estimated to be 190,000 years.

p	N
0.05	37,000
0.10	50,000
0.30	119,000
0.50	329,000

made for stream fishes. Araki *et al.* (2007) estimated effective population size for the entire population of steelhead trout (anadromous *Oncorhynchus mykiss)* in the Hood River, Oregon, at 1000 for the winter run and 1500 for the summer run. Sand darters are much more numerous that the steelhead. The data are not available to say which type of fish moves farther between natal site and breeding site; while steelhead commonly return to their natal stream to spawn, no data could be found about what percent return to the natal site without error or to within what distance of the natal site they return. No movement data are available for sand darters, but in an 18-month mark-and-recapture study, 1914 blackbanded darters *(Percina nigrofasciata)* were primarily recaptured within 33 m of their original capture site, with the longest recorded movement 420 m (Freeman, 1995). Given these data, it seems unlikely although not inconceivable that N for the sand darters may be 37,000 or higher, suggesting that lineage sorting is likely not responsible for the grouping of the MOB specimens with *A. bifascia* based on mitochondrial sequence.

Another line of evidence also suggests that lineage sorting issues may not be the reason that the nuclear data do not match the mitochondrial phylogeny. The nuclear estimate of divergence time between *A. beanii* and *A. bifascia* (t_1) was calculated to be 670,000 years, whereas the mitochondrial data (assuming a rate of evolution of 1.5% per million years in cytochrome *b)* indicate a divergence time of 490,000 years between the haplotype in western *A. beanii* and that in MOB *A. beanii*, a value only 73% as large. If lineage sorting issues caused MOB *A. beanii* to be more similar to *A. bifascia* than to western *A. beanii,* the divergence between the mitochondrial lineages within MOB *A. beanii* and *A. bifascia* would have to predate the population-level divergence between *A. beanii* and *A. bifascia* (see figure 2.9). Error margins on these estimates may be quite large (no way of estimating this error is known) so it is possible that their relative sizes may be incorrect. Still, the values of these estimates, along with the large effective population size required to make incorrect lineage sorting probable, suggest that lineage sorting issues were not responsible for the different pictures of relationships given by the mitochondrial versus the nuclear data.

Capture of mtDNA is another possible reason that the mitochondrial gene tree may not agree with the population level tree. Capture of mtDNA is characterized by selective introgression of mtDNA from one species into another following hybridization (even to the point of completely replacing the endogenous mtDNA); it has been documented (Avise, 1994; García-París *et al.,* 2003; Michel, 2005) though it is not common. For example, Wilson and Bernatchez (1998) found a population of

lake trout *(Salvelinus namaycush)* containing exclusively arctic char *(S. alpinus)* mitochondrial DNA, yet with apparently normal *S. namaycush* nuclear DNA. Dynamics of mtDNA capture are not well understood and probably vary case by case. It is possible that MOB *A. beanii* captured *A. bifascia* mtDNA through hybridization, but the *A. bifascia* mtDNA did not spread to western *A. beanii.* This scenario would explain why the estimated divergence between western and MOB *A. beanii* mitochondrial lineages does not predate the estimated nuclear divergence between *A. beanii* and *A. bifascia.* Phylogenetic analysis of a nuclear gene would reveal whether mtDNA has a different history than the nuclear genome, as would result from either lineage sorting or mtDNA capture.

It is also possible that the AFLP data do not reflect the true history of these groups. The AFLP analysis derives a distance measure between groups and does not distinguish primitive from derived states (indeed, this is impossible since there are no outgroup taxa in this analysis). It is possible that the groups have not been accruing genetic changes at the same rate. If *A. bifascia* has been changing genetically at a more rapid rate than MOB *A. beanii* and western *A. beanii*, the latter two could cluster in a distance analysis, even if the true history of the groups is that MOB *A. beanii* diverged most recently from *A. bifascia.* Lower polymorphism or heterozygosity in *A. bifascia* could potentially indicate some type of directional selection; indeed, *A. bifascia* has a percent of polymorphic loci lower than *A. beanii* (figure 2.5) and the lowest heterozygosity of the three groups (figure 2.5). The possibility of a higher rate of genetic change in *A. bifascia* cannot be ruled out,

although it seems extremely unlikely to affect the AFLP analysis since the entire nuclear genome is surveyed and the vast majority of markers are presumed to be from selectively neutral areas of noncoding DNA. Again, phylogenetic analysis of nuclear sequence data could resolve relationships between these groups and test whether this might be the case.

The current study is unable to definitively assess whether the AFLP or cytochrome *b* data reflect the true history. Without further information, the AFLP analysis is considered more likely to be a true indication of phylogenetic history even though it cannot distinguish primitive from derived states. This is because it is highly unlikely that the entire nuclear genome in *A. bifascia* is changing at a sufficiently elevated rate to cause MOB and western *A. beanii* to cluster if, as indicated by the phylogenetic analysis, *A. bifascia* is most closely related to MOB *A. beanii.* It seems more likely that the phylogeny of the mitochondrial cytochrome *b* gene does not reflect true taxonomic history due to evolutionary dynamics such as lineage sorting or mtDNA capture.

Conclusion

Affinity of the MOB A. beanii*, and evaluation of species status*

The MOB *A. beanii* specimens proved problematic in this study, with phylogenetic analysis of cytochrome *b* data indicating they should be placed within *A. bifascia,* and population genetic analysis of the AFLP data indicating that they are more similar to other *A. beanii* than to *A. bifascia,* but form a distinct group. As stated above, results of the AFLP analysis are favored as possibly indicating the

correct affinity of the MOB *A. beanii.* An interesting question is whether the current population genetic analysis would support species status for the MOB *A. beanii,* separate from western *A. beanii.*

Porter (1990) derived a population genetic method using F_{st} estimates to evaluate whether two distinct groups should be considered separate species. Although the population genetic analysis and the isolation by distance analysis demonstrate that MOB *A. beanii* is distinct from western *A. beanii,* Porter's method does not definitively say whether they should be considered separate species according to the population genetic data. Porter (1990) used gene flow estimates to test species boundaries where $Nm = (1/F_{st} - 1)/4$ is considered to be the number of migrants between groups per generation. (Slatkin (1985) derived another method for determining Nm based on frequency of private alleles, but his method is inappropriate for a data set containing only two demes and for data with a low-level error rate that makes low-frequency alleles especially suspect, such as is the case with AFLP.) The estimator of gene flow based on F_{st} depends on an island model of population structure with very restrictive assumptions such as an equilibrium under drift and migration and equal migration between all pairs of populations, among others. Porter, however, reasoned that calculation of gene flow between pairs of accepted sibling species (current gene flow assumed to be 0) related to the groups of interest could give an indication of the effect assumption violation generally has in the taxa in question. In Porter's method, gene flow estimates for related sibling species are compared to gene flow between the groups of interest to determine if there is similar

genetic isolation between the groups of interest as there is between accepted species. His related species comparison indicated that $0.01 \leq Nm \leq 0.15$ for sibling species in the group he was studying, and led him to conclude that for $Nm < 0.5$ two groups can be considered separate species, while with $Nm > 0.5$ other factors need to be looked at that could contribute to genetic similarity other than gene flow. For my analyses, Nm between western and MOB *A. beanii* is 1.09, greater than his 0.5 value, but Nm between the accepted species *A. beanii* and *A. bifascia* is 0.71, indicating that higher Nm obtains in this group between different species than in the group Porter looked at. Therefore, the 1.09 value is inconclusive according to Porter's method of analysis. In addition, population genetic sampling was sparse within the MOB for this type of analysis, with one population taken from the Alabama River (15 specimens) and one taken from the Tombigbee River (5 specimens). For a true evaluation using Porter's method, more populations would need to be added.

Comparing population genetic structure between groups

The data partition favored by the AFLP analysis, and considered the best current estimate of relationships within this group, is the division into *A. beanii* and *A. bifascia,* with a deep divergence found within *A. beanii* between populations in western drainages and populations in the MOB drainage. When population genetic parameters are compared between *A. beanii,* as the species is currently recognized, and *A. bifascia* (table 2.5), *A. beanii* has greater genetic diversity, larger average pairwise distance between populations, greater subdivision both at the level of drainage within species and at the level of population within drainage, and greater

Table 2.5. Comparison of population genetic parameters between the species *A. beanii* and *A. bifascia.* Ninety-five percent credible intervals shown for heterozygosity and θ^B.

Parameter	*A. beanii*	*A. bifascia*
Percent polymorphic loci	78.5; 75.8–81.6	57.4; 53.8–61.4
Genetic diversity measured by average heterozygosity	0.247; 0.241–0.252	0.177; 0.171–0.184
Pairwise distances between drainages and populations within drainages (in parentheses)	0.085 (0.043)	0.027 (0.026)
θ^B at the level of drainage within group and population within drainage (in parentheses)	0.153; 0.134–0.176 (0.061; 0.035–0.093)	0.074; 0.058–0.095 (0.041; 0.006–0.085)
Isolation by distance	r = 0.809 p = 0.001	r = 0.512 p = 0.060

genetic isolation by distance. All of these values are affected by the divergence between the western and MOB groups of *A. beanii,* however. Because this divergence (and the lack of a corresponding divergence in *A. bifascia)* is a recognizeddifference between the two species, the question arises as to how the species compare aside from this difference. Comparison, then, would be made between *A. bifascia* and each of the groups western *A. beanii* and MOB *A. beanii* separately. Such a comparison is shown in table 2.6. Western *A. beanii* shows the highest genetic diversity, followed by MOB *A. beanii*, and then *A. bifascia.* Genetic distance and θ^B are estimated both at the level of drainage within larger group and at the level of population within drainage for both western *A. beanii* and *A. bifascia,* but can only be estimated at the level of population within drainage for the MOB *A. beanii* (since it occupies a single drainage). Because only two populations were

Table 2.6. Comparison of population genetic parameters among the groups western *A. beanii,* MOB *A. beanii,* and *A. bifascia.* Ninety-five percent credible intervals shown for heterozygosity and θ^B.

Parameter	western *A. beanii*	MOB *A. beanii*	*A. bifascia*
Percent polymorphic loci	72.7; 69.1–76.7	47.5; 44.4–50.7	57.4; 53.8–61.4
Genetic diversity measured by heterozygosity	0.231; 0.225–0.237	0.204; 0.193–0.215	0.177; 0.171–0.184
Pairwise distances between drainages and populations within drainages (in parentheses)	0.047 (0.025)	N/A (0.061)	0.027 (0.026)
θ^B at the level of drainage within group and population within drainage (in parentheses)	0.075; 0.059–0.093 (0.022; 0.012–0.037)	N/A (0.099; 0.058–0.149)	0.074; 0.058–0.095 (0.041; 0.006–0.085)
Isolation by distance	r = 0.512 p = 0.060	N/A	r = 0.340 p = 0.020

sampled within the MOB drainage, one of which was represented by only six individuals, estimates of genetic distance and θ^B within the MOB are not reliable, and the best comparison is between western *A. beanii* and *A. bifascia.* Distance between drainages is higher in western *A. beanii* than in *A. bifascia* (though it is unknown whether this is significant), although distance between populations within drainages is the same in the two groups. θ^B is the same in both groups both at the level of drainage within larger group and at the level of population within drainage. Both groups show

some tendency to isolation by distance, although the correlation is only 0.340 in *A. bifascia* and the significance value only borders on significance ($p = 0.060$) in western *A. beanii.* These results suggest that the ancestor of *A. beanii* and *A. bifascia* had similar genetic distance between drainages, genetic subdivision, and tendency toward isolation, and that these population genetic characteristics were passed to the daughter species through the speciation event unchanged. It cannot be inferred whether genetic diversity changed during or subsequent to the speciation event, or which group changed from the ancestral condition. Population genetic analysis of an outgroup could possibly address the question of the direction of change, but there is no single species that is the closest relative to the *A. beanii/A. bifascia* species pair, rather they are most closely related to the species pair *A. meridiana/A. vivax* (Near *et al.,* 2000; Wiley and Hagen, 1997; Simons, 1992).

Chapter 3: Examination of population genetic characteristics in *Fundulus escambia* and *F. nottii*

The goal of this part of my work was to quantify population genetic parameters of a pair of sister species of starhead topminnows *(Fundulus nottii* and *F. escambiae)* that speciated within the last four million years (Wiley, 1977). Characterization of population genetic structure of closely related species enables inferences to be made about the genetic structure of the ancestor. There has not previously been any assessment of the population genetics of any species in the *Fundulus nottii* species group.

Introduction

Rationale

Do speciation events leave a recognizable population genetic pattern in the species formed? To begin addressing this question, a sister species pair, *Fundulus nottii* and *F. escambiae,* was identified that has three key characteristics: (1) the date of the speciation event was approximately known and was recent enough to expect a population genetic pattern to still be evident, (2) evidence suggested there is no third species, extinct or extant, that is more closely related to either species of this pair than they are to each other (Smith, 1981; Wiley and Mayden, 1985), and (3) the occurrence of each species within several river drainages allows clear demarcation of population subdivisions for population structure analysis. In addition, specimens were available in the KU Natural History Museum tissue collection.

Sequence information from the mitochondrial cytochrome *b* gene was chosen for the phylogenetic analysis because this gene is straightforward to sequence and has

an appropriate amount of variation to address relationships at this level. AFLP, or Amplified Fragment Length Polymorphism, was chosen to study population genetic parameters because it yields a large number of genetic markers, has been shown to be highly reliable, surveys the entire genome, and requires no prior knowledge of genome sequences.

Study organism

Fundulus nottii (known by various common names including southern starhead topminnow) and *F. escambiae* (known as the russetfin topminnow or eastern starhead topminnow) compose a species pair living in fresh to slightly brackish water in the southeastern United States. Both species mature at a maximum size of 65 mm SL (Ross, 2001; Boschung and Mayden, 2004). They prefer stream margins and backwaters with no to slight current over sand or muddy bottoms. They tend to forage and swim near the surface and high in the water column. Although *F. nottii* was described in 1854 by Agassiz (called *Zygonectes nottii)* and *F. escambiae* in 1886 by Bollman (as *Z. escambiae),* the taxonomic history of the species is complex. Along with several other species, Garman synonymized *Z. escambiae* with *Z. nottii* in 1895. Jordan and Everman (1896) made *Zygonectes* a junior synonym of *Fundulus,* but still considered *F. escambiae* to be a synonym of *F. nottii.* These species were not recognized in their current delineation until the work of Wiley in 1977. *F. nottii* and *F. escambiae* occupy adjacent, non-overlapping, fairly restricted ranges (figure 3.1), with *F. nottii* found on the coastal plain within the Mobile Bay drainage (MOB) and drainages to the west, and *F. escambiae* on the coastal plain in drainages

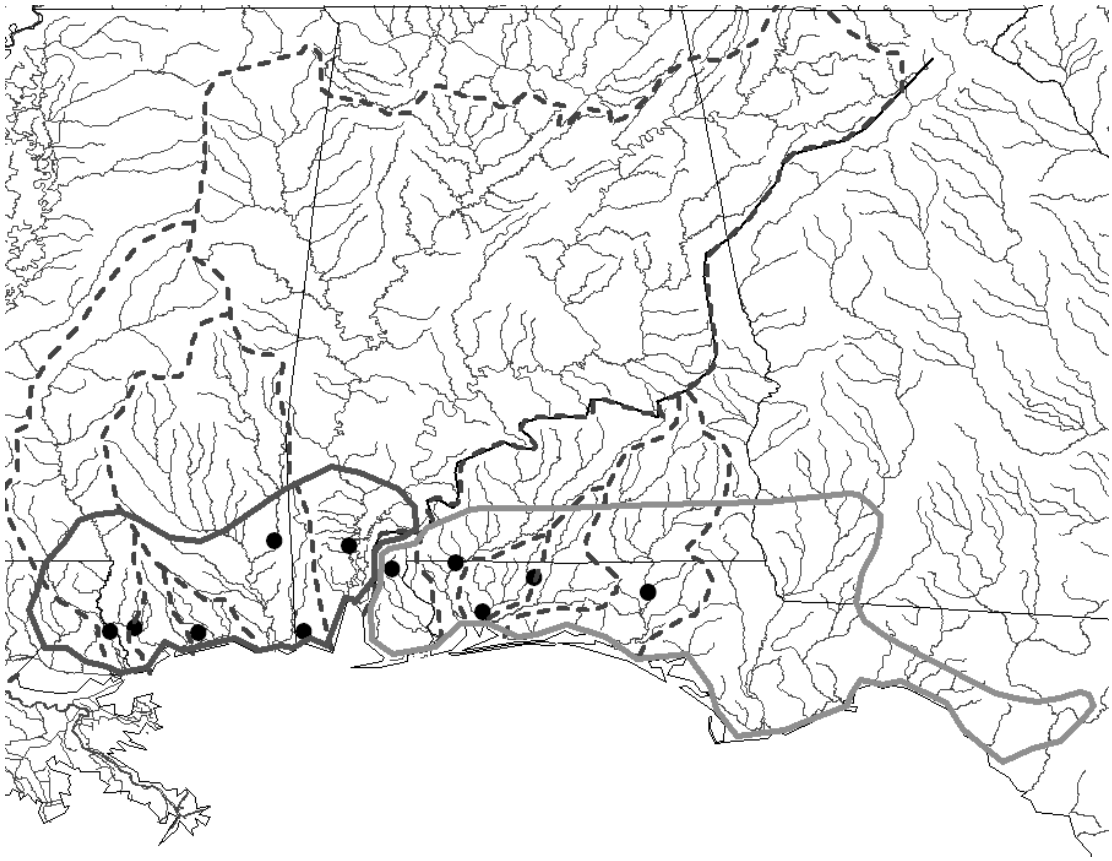


Figure 3.1. Ranges of *F. nottii* (west, dark line) and *F. escambiae* (east, grey line). Ranges are represented by solid lines with drainages shown by dashed lines for reference.

immediately to the east of MOB. There are five recognized species within the *F. nottii* (or starhead topminnow) species complex, which together are placed within the subgenus *Zygonectes*. Evidence suggested that *F. nottii* and *F. escambiae* are closest sister species and that neither had undergone a more recent speciation with subsequent loss of the other species formed. Wiley and Mayden (1985) point out that the continuous distribution of the *F. nottii* species group leaves no reason to think that extinctions have occurred in the group. While Cashner *et al.* (1992) postulated a sister relationship between *F. nottii* and *F. lineolatus* based on allozyme data, both

morphological (Wiley, 1977) and molecular data (Ghedotti and Grose, 1997) find a sister relationship between *F. nottii* and *F. escambiae* with strong support.

Methods and materials

Experimental overview

For this project, two different types of molecular data were used: cytochrome *b* gene sequence, and AFLP (or Amplified Fragment Length Polymorphism) fingerprints. The sequence data were used to verify that two distinct groups (species) were present and that the geographic boundary between the groups was correctly identified, and to estimate genetic distance between the species. The AFLP fingerprints were used for population genetic analysis.

Samples

All samples were collected in the 1980s and are housed in the University of Kansas Natural History Museum tissue collection (see Appendix 3). Specimens were frozen in liquid nitrogen when collected and transferred to –80º C for long-term storage. Most of the *Fundulus* specimens used in this study were stored in plastic bags in –80º C and accessioned into the tissue collection at the time of this study (tissue numbers > 6558). Some of the specimens from three different collecting events were previously accessioned into the tissue collection (tissues 633, 525, 542, 629, 901, 902, and 1621). A total of 20 *F. nottii* specimens, 12 *F. escambiae* specimens, and three outgroup specimens (2 *F. lineolatus,* and 1 *F. dispar*) were sequenced (Appendix 4). See figure 3.2 for collection locations and numbers

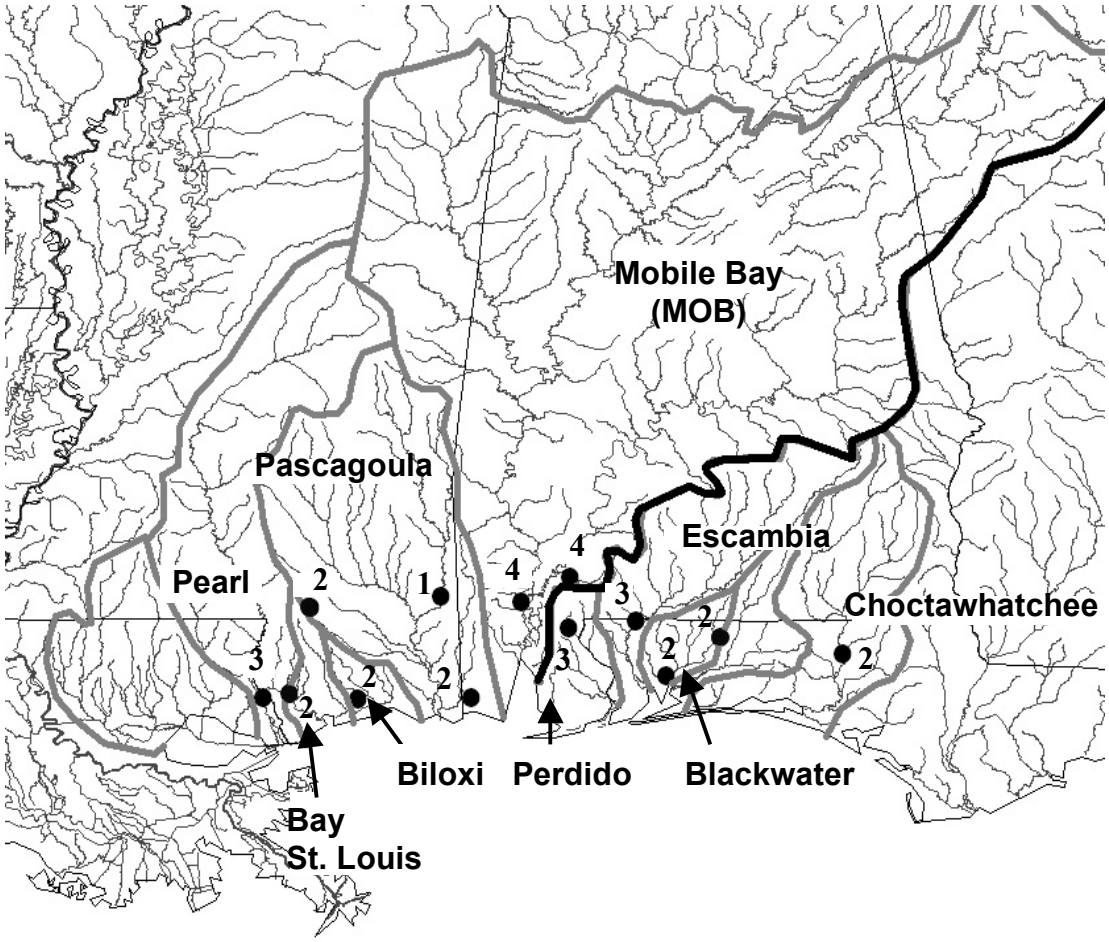


Figure 3.2. Specimens used in the phylogenetic analysis of cytochrome *b* sequence. Black dots indicate collection sites; number of specimens used from each site is indicated. Grey lines indicate drainage boundaries. Black line indicates boundary thought to demarcate the ranges of the two species.

sampled for the sequencing work. For the AFLP work, samples were chosen from as wide a geographic distribution as possible from within each species range to get the best possible estimate of genetic diversity within species (figure 3.3). An average of 15 specimens were used from each population, with a range in numbers used of 12 to 20. Six populations were used for *F. nottii*, representing five drainages and consisting of 88 individuals. Five populations representing four drainages (77 individuals) were

used for *F. escambiae.* For each of these species, one drainage has more than one population sampled so that genetic structure could be estimated within drainages as well as among them.

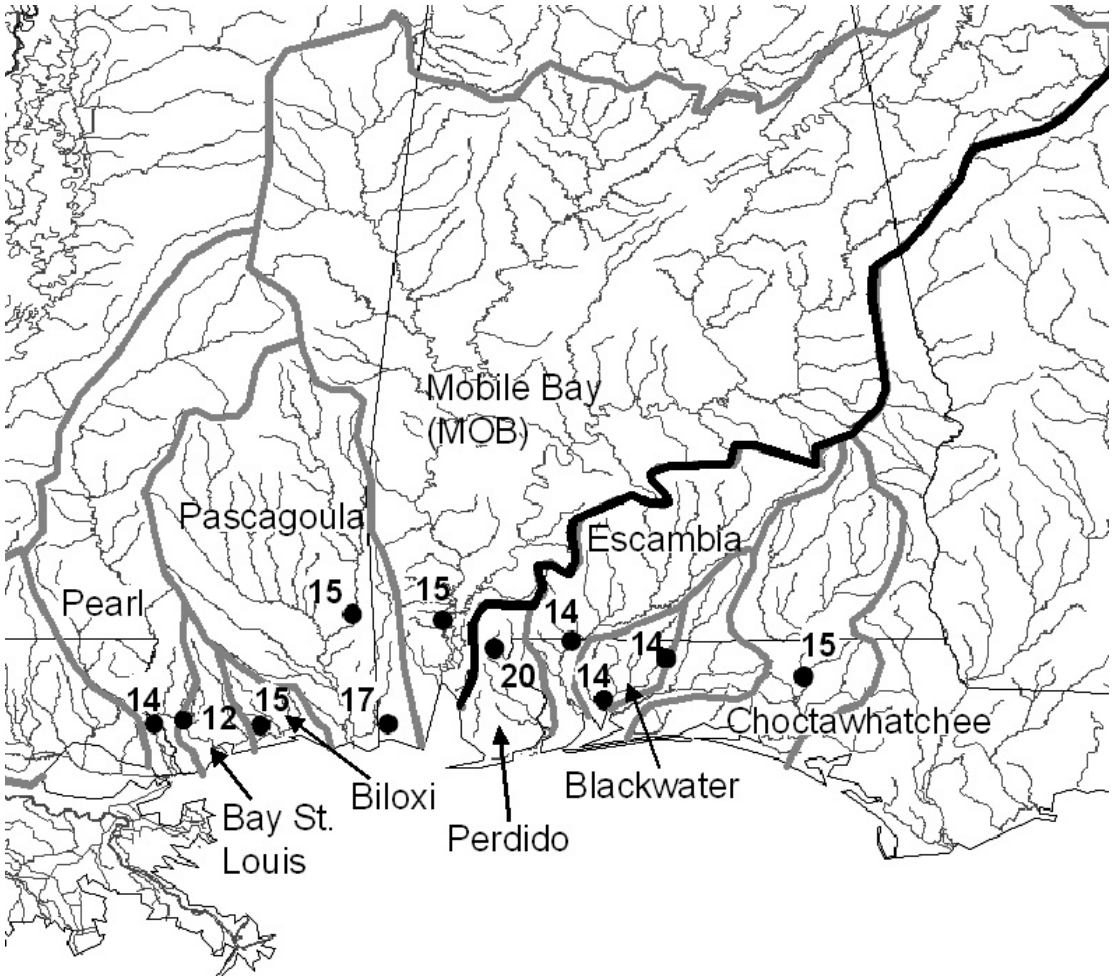


Figure 3.3. Drainages represented by samples of *Fundulus* in the AFLP work. Number of specimens included is shown near each collection site. The black line separates the range of *Fundulus nottii* (west) from that of *F. escambiae* (east).

Molecular methods – sequencing

Nearly all (1131 of the 1140 bases) of the mitochondrial gene cytochrome *b* was sequenced. Nucleic acids were extracted with Qiagen DNeasy DNA extraction kits, following company protocols. Extractions were then quantified on a NanoDrop ND-1000 spectrophotometer. We used the cytochrome *b* primers given in Song *et al.* (1998), and their PCR protocol. PCR was performed with puReTaq Ready-To-Go PCR beads (Amersham Biosciences Corp, Piscataway NJ). Reactions contained 1.0 μL of each of the forward and reverse primers (10 pmol dilution), 2.5 units of DNA polymerase, 200 μM of each dNTP in 10 mM Tris-HCl, 50 mM KCl, and 1.5 mM $MgCl_2$ plus enough DNA sample to yield 100 ng in a total volume of 25 μL. Amplicons were purified enzymatically using ExoSap-IT (USB Corp. Cleveland, OH). We used 1/5 the manufacturer-suggested amount of enzymes for purification and incubated samples for 31 minutes at 37˚ C, followed by 15 minutes at 80˚ C to inactive the enzymes. The gene was sequenced on an ABI 3130xl sequencer using both forward and reverse primers. The program Sequencher 4.1.1 (Gene Codes Corp.) was used to align forward and reverse reads, edit ambiguous base readings, and produce consensus sequences. Alignment of the protein coding sequences was straightforward and sequences were translated into amino acids in the application Se-Al v1.0a1 to verify alignment.

Phylogenetic analysis of sequence

Parsimony analysis was run on the cytochrome *b* sequence data. A heuristic search using 10,000 random stepwise-addition sequences was performed in PAUP*

v4.0b10 (Swofford, 2003). Tree support was assessed from 10,000 bootstrap replicates.

Molecular methods – AFLP analysis

DNA was extracted with Qiagen DNeasy Tissue Kits and quantified on a NanoDrop ND-1000 Spectrophotometer. LI-COR AFLP protocols were followed with no modification. In the basic AFLP process, genomic DNA is digested with *Eco*RI (a 'rare cutter' with a 6-base recognition sequence) and *Mse*1 (a 'frequent cutter' with a 4-base recognition sequence), producing millions of fragments. Because *Mse*1 cuts much more frequently than *Eco*RI, more than 90% of fragments have two *Mse*1 ends, and most of the remaining fragments have one *Eco*RI end and one *Mse*1 end. After restriction of the DNA, short known DNA sequences or "adaptors" are ligated to the fragment ends, one for ends left by *Eco*RI cutting and another for ends cut by *Mse*1. The fragments are then amplified with selective primers complementary to the cut sites and adaptors plus one additional base (A for the *Eco*RI end primer and C for the *Mse*1 end primer) in a 'preamplification' step. This step amplifies 1/16 of the fragments originally present (*i.e.*, only those that contain the specific base on each end that matches the additional base used in the primer). Fragments with two *Mse*1 ends are lost (suggested explanations given in Vos, 1995), resulting in almost exclusive amplification of fragments with one *Eco*RI end and one *Mse*1 end. Next, in the 'selective amplification,' pairs of primers are used such that one primer of each pair is complementary to the *Eco*RI cut site, adaptor, and added A plus two additional selective bases, and the other is

complementary to the *Mse*1 cut site, adaptor, and added C plus two additional selective bases. Each selective primer is designated according to the type of cut (*Eco*RI or *Mse*1) to which it is complementary plus the identity of the three additional selective bases. For instance, a primer complementary to an *Mse*1 cut (and adaptor) plus the added C, plus the additional selective bases A and T would be called *Mse*1-CAT. The LI-COR kit provides eight different options for the two additional selective bases on the *Mse*1-end primer and eight for the *Eco*RI-end primer, yielding 64 different possible primer pairs. Each primer pair produces a unique set of fragments. *Eco*RI primers are fluorescently labeled so that resulting fragments are also labeled. Samples were run on 25 cm gels (6.5% KB + acrylamide), with disposable 64-well shark tooth comb. The gel was pre-heated for 25 minutes at the following conditions: 1500v, 40W, 45mA and 45º C. Samples were denatured, then 0.85 μL of sample was loaded in wells and run for 3 hours 10 minutes at same conditions as pre-heat. In this study, fragments were visualized on a LI-COR Long Readir 4200 DNA sequencer, which uses a laser to detect fragments and sends a digital image to a connected computer. Two different wavelengths of dye are used, allowing sets of fragments from two different primer pairs to be visualized independently (one on each wavelength) on each gel. Primer pairs were screened using three individuals of each species until four primer pairs were found that revealed polymorphic fragments: *Eco*RI-AAC/*Mse*1-CAT, *Eco*RI-ACT/ *Mse*1-CAC, *Eco*RI-ACA/ *Mse*1-CAT, *Eco*RI-ACT/ *Mse*1-CTC. A total of ten gels were scored

for this project, each with bands on two wavelengths, yielding 20 gels worth of data (5 for each primer pair).

Distance analysis of sequence

Kimura 2-parameter genetic distance was derived for every possible pairing of an *F. escambiae* with an *F. nottii* sequence in the program Phylip (Felsenstein, 2004). The average is reported as a measure of the genetic distance between the species.

Gel scoring

"Loci" were identified by primer and length of amplified fragment, and each fragment length was considered to be a unique locus. Therefore, for example, *Eco*RI-AAG/*Mse*1-CTA 337 is assumed to designate a unique location in the genome that produces a fragment 337 base pairs long when AFLP amplification is performed with the primer pair *Eco*RI-AAG and *Mse*1-CTA. AFLP bands behave as dominant genetic markers, where presence of a band indicates at least one copy of the relevant DNA and dominant homozygotes cannot be distinguished from heterozygotes. Absence of a specified band is interpreted as an individual being homozygous for the (recessive) genotype that does not produce the fragment in question. All interpretable loci (bands not too faint or close to another locus) were scored using LI-COR Saga Automated AFLP Analysis Software and rechecked manually. Presence of a band was scored as '1' and absence as '0'. Very faint or indistinct bands within loci were scored as '?'. For a given individual and primer pair, there is no reason to think that bands of consecutive length on the gel are in any way associated biologically. Therefore, it was considered extremely unlikely that an individual should be missing

three or more consecutive bands present with high frequency in other individuals. Individuals lacking such consecutive bands were re-run through the entire AFLP process due to the possibility of incomplete digestion of DNA. In addition, 16 *Fundulus* samples were rerun randomly as a means to assess the repeatability of the method.

AFLP data analysis

The following assessments were made on the AFLP data: descriptive fragment analysis (including percent polymorphic loci, possible group-diagnostic loci, and private alleles), isolation by distance analysis, genetic diversity measured by average expected heterozygosity (Weir, 1996), genetic distance between groups (Nei, 1978), and population subdivision measured by measured as θ^B (Holsinger *et al.,* 2002), an analog of Wright's F_{st}. In addition, an assessment was made of repeatability of the method.

Percent polymorphic loci (and a 95% confidence interval based on 10,000 bootstrap replicates) was estimated in an Excel macro written by the author, and was evaluated at the 95% level, where a locus is considered polymorphic if the less common allele is present at a frequency of at least 5%. The Lynch and Milligan (1994) Taylor expansion method for unbiased estimation of allele frequencies from dominant marker data was used. Possible species-diagnostic loci and private alleles were identified by eye using allele frequency output for each locus from the program Hickory (Holsinger *et al.,* 2002)**.** Possible group-diagnostic loci are loci in which one group is completely missing an allele while the other group is fixed for that allele,

while private alleles are alleles present in only one group. Isolation by distance analysis was conducted using a Mantel test in the software package TFPGA (Miller, 1997) with 999 permutations. Isolation by distance analysis compares a matrix of genetic distances between populations with a matrix of geographical distances between the localities where the populations were collected to see whether migration primarily happens between adjacent populations (genetic similarity greatest for adjacent populations). Average heterozygosity, a measure of the expected frequency of heterozygotes, and θ^B, a measure of the evenness of allele distribution among subgroups within a population were derived in Hickory (Holsinger *et al.,* 2002), a program specifically designed for the analysis of dominant genetic data by a Bayesian method. Credible intervals at the 95% level are reported for each of these statistics. Default values for burn in (5,000) and sampling (25,000) were used, as were all default parameter values. In each analysis, the model assuming no within-population inbreeding (f = 0) was preferred, both because of model fit and because of lack of biological reason to suspect inbreeding. In its current implementation, Hickory cannot perform nested analyses. Nei's (1978) unbiased measure of genetic distance between groups was estimated in the software package PopGene 1.32 (Yeh *et al.,* 1997). Values reported are average pairwise distances. No way of testing significant difference between these values is known, both because individual comparisons are nonindependent and because genetic distance is correlated with geographic distance. In all analyses of population differentiation, within drainage estimates are based on

two populations within the Pascagoula River drainage for *F. nottii* and two populations within the Blackwater River drainage for *F. escambiae.*

To assess method repeatability, 16 individuals were selected at random and rerun through the AFLP process. Scores obtained on repeated runs of the same individual were compared and rate at which scores differed for the same individual between runs was reported.

Results

Phylogenetic analysis

The 1131 bases of the cytochrome *b* gene from 35 individuals of *Fundulus* contained 258 variable sites (32 in first positions, 8 in second positions, and 218 in third positions) of which 175 sites were phylogenetically informative (19 in first, 1 in second, and 155 in third positions respectively). There was evidence for base composition bias across all three codon positions (0.240 A, 0.332 T, 0.273 C, 0.155 G), which was more pronounced in third positions (0.304 A, 0.300 T, 0.334 C, 0.061 G). Plots of Tamura-Nei distance versus total transitions and transversions for all positions (data not shown) showed no saturation for either transitions or transversions in any codon position.

As can be seen in figure 3.4, the parsimony analysis supported the division of the specimens into two species across the eastern boundary of the Mobile Bay Drainage, recovering a clade of all *F. nottii* with 100% bootstrap support and a clade of all *F. escambiae* with 99% bootstrap support. Individuals grouped together by

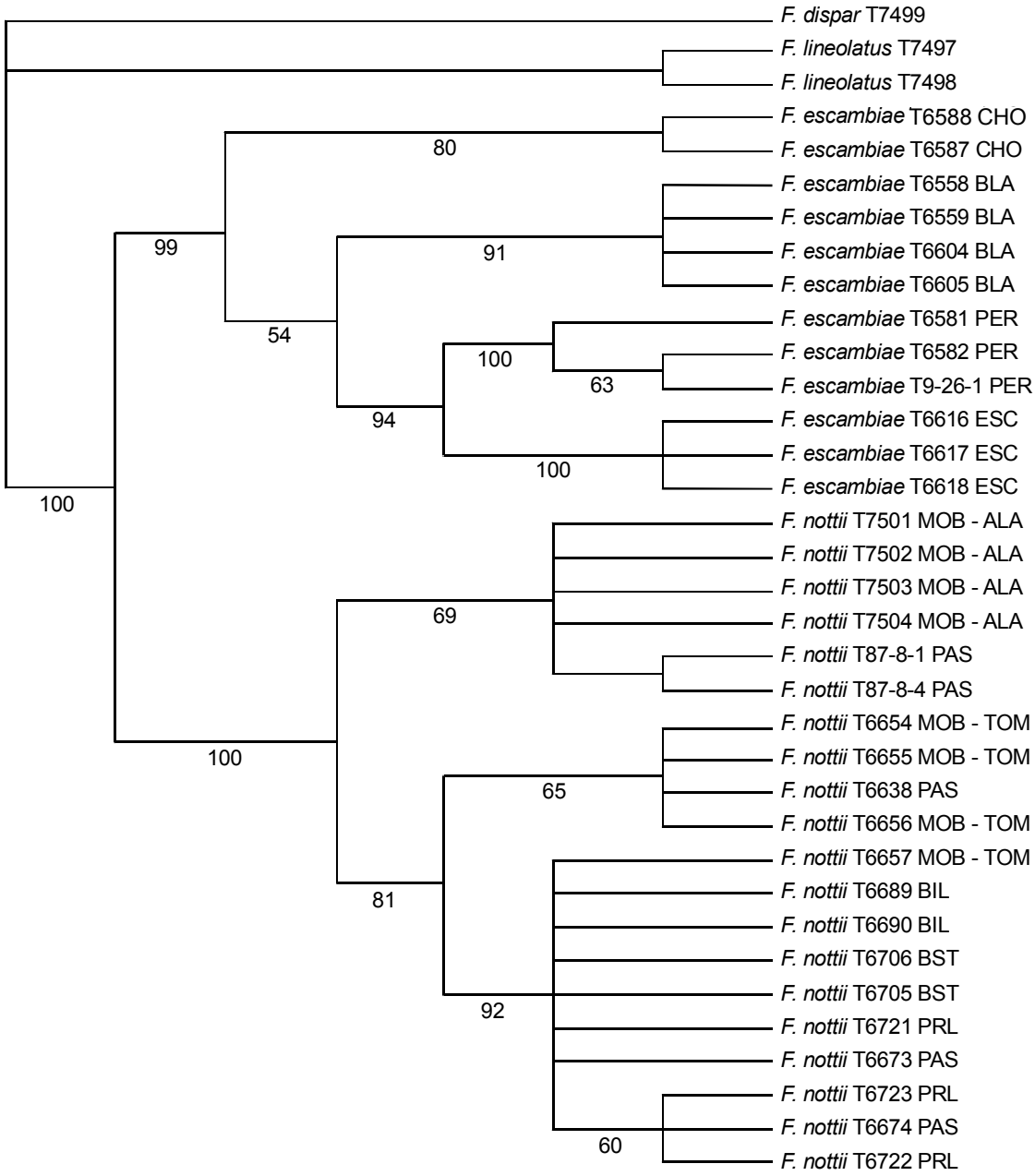


Figure 3.4. Phylogenetic tree based on cytochrome *b* mitochondrial sequence. Parsimony bootstrap values (10,000 replicates) are placed below nodes. The sequence data was analyzed by parsimony in PAUP* (Swofford, 2003). Drainages from which individuals were used were Pearl (PRL), Bay St. Louis (BST), Biloxi (BIL), Pascagoula (PAS), Mobile Bay (MOB), Perdido (PER), Escambia (ESC), Blackwater (BLA), and Choctawhatchee. Drainage abbreviations are as used in Swift *et al.,* 1986, with the exception that 'MOB – TOM' means from within the Tombigbee R. in the MOB drainage and 'MOB – ALA' means from within the Alabama R. in the MOB drainage.

drainage within the *F. escambiae* clade, but there was less separation by drainage among the *F. nottii* specimens where several clades of moderate support grouped specimens from different drainages.

Distance analysis of sequence

Average pairwise Kimura 2-parameter distance between cytochrome *b* sequences of *F. nottii* individuals and *F. escambiae* individuals was 0.0628.

Fragment analysis – polymorphic loci, diagnostic loci, private alleles

A total of 190 loci were scored, of which 11 were completely monomorphic and none had rare allele frequency greater than 5% making 94.2% of loci polymorphic across the entire data set using the 95% criterion. Within *F. escambiae,* 61.1% (95% CI: 57.4–65.2) of loci were polymorphic; within *F. nottii*, 68.9% (95% CI: 64.7–73.2). There were no loci diagnostic for species, although nine loci came close to being diagnostic (table 3.1). Each species had a total of ten private alleles. In addition, there were many loci for which an allele was present in greater than 50% of individuals of one species but in two or fewer individuals of the other species: seven for *F. escambiae* and 25 for *F. nottii.*

Isolation by distance analysis

Results of the isolation by distance analysis are shown in figure 3.5. As shown by the Mantel test, correlation between genetic and geographical distance is not significant for either *F. escambiae* or *F. nottii* populations, although correlation is nearly significant for *F. escambiae* ($p = 0.051$) with correlation coefficient 0.631.

Table 3.1. Fragments nearly diagnostic for species. Shown are proportions of individuals for which the band is present.

#	Fragment	*F. escambiae*	*F. nottii*
1	*Eco*RI-AAG/*Mse*1-CTA 478	1/76	87/88
2	*Eco*RI-AAG/*Mse*1-CTA 262	0/76	85/87
3	*Eco*RI-AAG/*Mse*1-CTA 163	1/75	86/88
4	*Eco*RI-ACT/*Mse*1-CTT 185	2/76	85/86
5	*Eco*RI-ACT/*Mse*1-CTT 169	76/76	1/88
6	*Eco*RI-ACT/*Mse*1-CTT 145	2/73	86/87
7	*Eco*RI-ACT/*Mse*1-CTG 412	74/76	0/86
8	*Eco*RI-ACT/*Mse*1-CTG 244	0/76	85/86
9	*Eco*RI-ACT/*Mse*1-CTG 148	2/75	88/88

Population genetic analysis – genetic diversity

At the species level, *F. escambiae* heterozygosity is significantly lower than that of *F. nottii* (0.209 versus 0.241, respectively). Table 3.2 shows heterozygosities at the species, drainage and population level. Heterozygosity was only derived at the drainage level for two drainages, the Pascagoula River drainage (0.210) for *F. nottii* and the Blackwater River drainage (0.138) for *F. escambiae*. Average population heterozygosity was 0.167 (range 0.140 to 0.184) for *F. nottii* and 0.144 (range 0.095 to 0.184) for *F. escambiae.* When the outlier population BLA1 was removed, average population heterozygosity was 0.157 for *F. escambiae.*

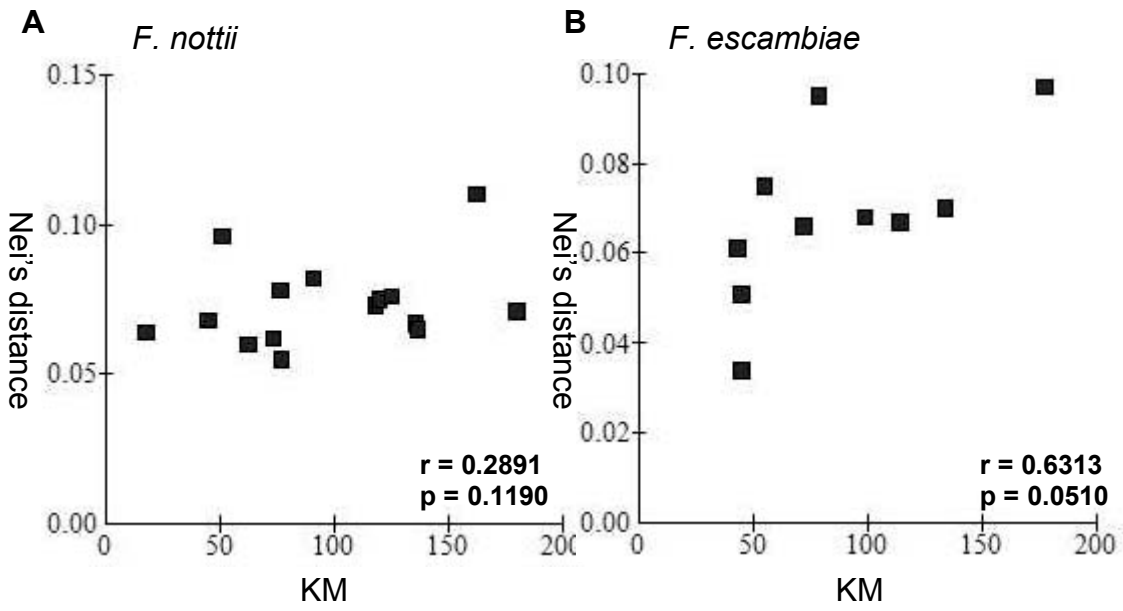


Figure 3.5. Plot of Nei's genetic distance (pairwise comparisons of all individuals in populations) versus geographical distance between populations as assessed in a Mantel test (A) for *Fundulus nottii*; (B) for *F. escambiae*.

Population genetic analysis – genetic distance

Figure 3.6 shows Nei's (1978) average unbiased pairwise genetic distance between populations across species, as well as the average and range of pairwise distances between drainages within each species, and distance between populations within the same drainage for each species. Average distance between populations selected from different species was 0.562 (range 0.482–0.636), while average distance between *F. nottii* drainages was 0.066 (range 0.050–0.075) and that between *F. escambiae* drainages was 0.063 (range 0.047–0.093). Two populations from within the Pascagoula River drainage were available to estimate within-drainage distance between populations for *F. nottii* and the distance between these populations

Table 3.2. Heterozygosity (H) at the population and drainage level. Drainage-level heterozygosity was only assessed for one drainage per species. Ranges represent 95% credible intervals. Average population heterozygosity was 0.167 for *F. nottii* and 0.144 for *F. escambiae.* When the outlier population BLA1 was removed, heterozygosity was 0.217 (0.210–0.225) for *F. escambiae.* Heterozygosities were calculated in the program Hickory (Holsinger *et al.,* 2002). A t test between the species-level heterozygosities is not significant ($t = 1.14$, $p > 0.2$).

Species	Species-level H	Drainage	Drainage-level H	Population	Population-level H
F. nottii	0.241 0.235–0.248	PRL		PRL	0.171 0.157–0.185
		BST		BST	0.176 0.163–0.189
		BIL		BIL	0.163 0.150–0.176
		PAS	0.210 0.200–0.219	PAS1	0.184 0.171–0.196
				PAS2	0.173 0.161–0.186
		MOB		MOB	0.140 0.127–0.154
F. escambiae	0.209 0.202–0.216	PER		PER	0.139 0.128–0.151
		ESC		ESC	0.171 0.159–0.184
		BLA	0.138 0.127–0.149	BLA1	0.095 0.082–0.108
				BLA2	0.133 0.121–0.147
		CHO		CHO	0.184 0.173–0.196

was 0.050. Two populations from within the Blackwater River drainage were used to derive an estimate for *F. escambiae* and the distance between them was 0.030. Because MOB populations were very distinct from populations to the west in the *Ammocrypta* work (chapter 2), assessment was also made of distance between the

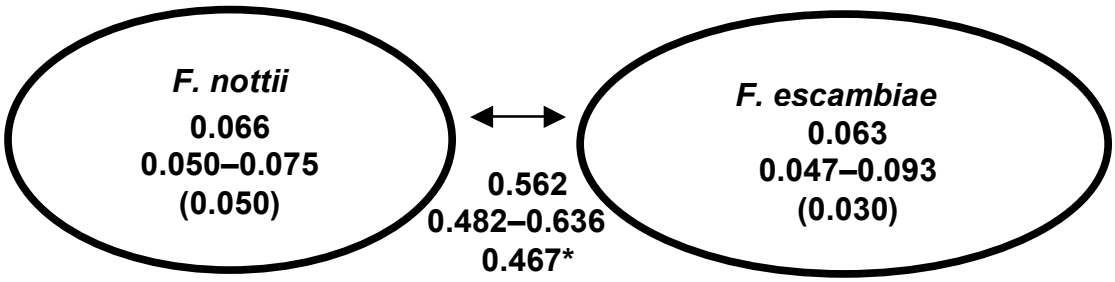


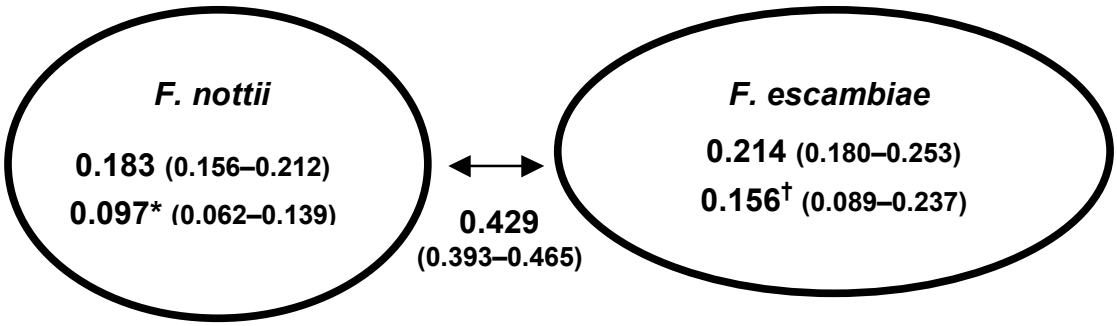
Figure 3.6. Nei's (1978) unbiased genetic distances between groups calculated in PopGene (Yeh *et al.,* 1997). Upper numbers are average pairwise distances between drainages either within species (inside circles) or between species (in center under arrow). Ranges in pairwise distances between drainages are also given. Numbers in parentheses are average pairwise distance between populations within drainages (measured for one drainage per species). Starred number is distance between species when all individuals within species are treated as one group instead of being divided into populations (given for comparison with cytochrome *b* sequence distance).

MOB population and other *F. nottii* populations. The average distance was 0.081 (range 0.060 to 0.105). When individuals in each species were clustered into one group (not treated as separate populations) Nei distance between the two species was 0.467. This number is presented for later comparison with sequence divergence, which is derived without division of individuals into populations.

Population genetic analysis – θ^B

Figure 3.7 shows genetic differentiation (measured as θ^B, an analog of Wright's F_{st}) at three different levels of analysis: between the two species, among drainages within each species, and at the level of population within drainage for each species. Differentiation is great between the two species (0.429) and also high at the level of drainage within species (0.183 for *F. nottii* and 0.214 for *F. escambiae)*. It is somewhat lower at the level of population within drainage (0.097 within the

Pascagoula River drainage for *F. nottii* and 0.156 within the Blackwater River drainage for *F. escambiae)*, although the decrease at the within- versus between-drainage level is only significant in the case of *F. nottii.*



*** Based on Pascagoula River drainage**
† Based on Blackwater River drainage

Figure 3.7. θ^B at different levels of analysis. θ^B at species level shown under arrow between species (represented by circles). Within species, upper number is drainage within species and lower number is population within drainage. Numbers in parentheses are 95% credible intervals. θ^B was calculated in the program Hickory (Holsinger *et al.,* 2002).

Error analysis

Both runs were successful for 13 of the 16 individuals rerun through the AFLP process. For these 13 individuals, 6.1% of scores (of a total of 2363 scores) differed between runs.

Discussion

Phylogenetic analysis

Phylogenetic analysis of the cytochrome *b* data supports division of these specimens into previously recognized species (two species divided across the eastern boundary of the MOB). *F. escambiae* grouped into clades by drainage while *F. nottii*

did not, perhaps indicating lower rate of migration between drainages in *F. escambiae* than in *F. nottii*.

Distance analysis of sequence

The amount of cytochrome *b* sequence divergence is not incompatible with a speciation event between 2 and 4 million years ago. A divergence of 0.0628 in cytochrome *b* sequence between *F. nottii* and *F. escambiae* implies that each species is 0.0314 dissimilar from the common ancestor. The vicariant event during which they are believed to have diverged is thought to have taken place within the last 2–4 million years (Wiley, 1977). It is generally thought that the cytochrome *b* gene evolves at a rate of 1 to 2 percent per million years in poikilotherms (Thorp *et al.,* 2005), although the exact rate varies by taxonomic group, seemingly with some correlation to body size, temperature, and even clutch size (Bromham, 2002; Gillooly *et al.,* 2004). Poulkakis *et al.* (2005) estimated the rate of cytochrome *b* evolution to be 1.55% per million years in lizards of the genus *Podarcis,* while Guiking *et al.* (2006) found a general mitochondrial evolution rate of 1.35% per million years for *Natrix* snakes. Domingues *et al.* (2005) found a higher rate of cytochrome *b* evolution at 2.5% per million years in *Chromis* fishes, however. Given that different taxa exhibit different rates of sequence evolution, it is not possible to use sequence divergence to pinpoint a time of speciation without some way to calibrate rate of evolution for the taxon in question. However, based on the studies above, it is probable that *Fundulus* cytochrome *b* evolves at a rate falling within the range of

1.5% to 2.5% per million years. Such a range would place the speciation event somewhere between 1.3 and 2.5 million years ago.

Fragment analysis

'Fragment analysis' included descriptive measures on the AFLP fragments such as percent polymorphic loci and identification of group-diagnostic loci. Based on the fragment analyses, these species have a high level of polymorphism and are very distinct from each other. The much higher polymorphism level (93.7%) across both groups than within either group individually (61.1% and 68.9%) indicates that the groups are fixed or nearly fixed for different alleles at many loci. In addition, the presence of ten private alleles in each group and the many loci with high frequency in one group and near absence in the other indicates substantial genetic distinctness between the species.

Because of the error rate in the AFLP analysis, it is unlikely that an absolutely group-diagnostic locus could be detected by this method of analysis. While no absolutely diagnostic loci were found, the nine nearly group-diagnostic bands listed in table 3.1 allow a probability greater than 99.99% of correctly assigning unknown individuals to the correct species. This is based on proportion of individuals for which group identity is correctly indicated by each fragment and assuming an individual will be assigned to a group whenever the majority of bands so indicate. Group identity is considered to be 'correctly indicated' when an individual possesses the character state most common for the group to which it belongs. Since nine fragments are nearly diagnostic between *F. nottii* and *F. escambiae*, the probability of

correctly assigning between these groups is equal to the proportion of all individuals for which at least five fragments are expected to correctly indicate group.

Isolation by distance analysis

F. escambiae shows a nearly significant amount of isolation by distance and fairly high correlation coefficient, while *F. nottii* shows much less of a pattern of isolation by distance. This may indicate that *F. escambiae* has less geographically widespread gene flow than *F. nottii.*

Population genetic analysis – genetic diversity

Species-level heterozygosity is significantly higher in *F. nottii* than *F. escambiae.* Drainage-level heterozygosity appears to be higher for *F. nottii,* but only one drainage was assessed per species. Unfortunately, the drainage chosen for *F. nottii* contained the populations with the highest heterozygosity for that species, while the drainage chosen for *F. escambiae* contained the two lowest-heterozygosity populations, making the comparison of questionable validity. Average population heterozygosity was higher in *F. nottii* than in *F. escambiae*.

Population genetic analysis – genetic distance

Nei's (1978) genetic distance between drainages was similar in both *F. nottii* and *F. escambiae* and much smaller than the distance between the species, reinforcing the phylogenetic analysis of sequence in indicating that *F. nottii* and *F. escambiae* are distinct groups. The small distance between MOB populations and the rest of *F. nottii* populations also supports the phylogenetic analysis of sequence data in indicating that MOB populations belong in the species *F. nottii.* An estimate of

divergence time can be derived from Nei's distance. Nei's distance is approximately $2\mu t$ where μ is the infinite alleles mutation rate at the loci examined. Assuming a species divergence time from 1.3 to 2.5 million years ago as per the sequence divergence calculations, the between-species value of 0.562 for Nei's distance would yield a mutation rate of 1.8×10^{-7} to 9.3×10^{-8} mutations per year for these AFLP loci. Assuming that these mutations are allele loss and given that any one of ten DNA sites may change to prevent a fragment from being recovered (6 in the *Eco* RI recognition site and 4 in the Mse 1 recognition site), this translates to a mutation rate of 1.8×10^{-8} to 9.3×10^{-9} changes per nucleotide site per year (or generation, since the generation time for these fishes is one year). This is slightly larger than the genomic mutation rate estimates of Drake *et al.* (1998), even when these estimates are corrected to represent mutation rate per generation. This correction is done by multiplying their estimated mutation rate per cell division by the estimated number of divisions involved in gamete production. Their numbers, once this correction is made, are 8.5×10^{-9} mutations per base per generation for *Drosophila melanogaster* and 2.0×10^{-9} mutations per base per generation for *Caenorhabditis elegans*. It is on the same order, however, and not unbelievable, particularly for the larger divergence time of 2.5 million years (which yields the lower estimate for mutation rate). Therefore divergence estimates derived from different genomes (nuclear versus mitochondrial) and by different methods (AFLP marker-based versus sequence-based) are mutually compatible.

Population genetic analysis – θ^B

There is considerable genetic differentiation among drainages within *F. nottii* and *F. escambiae* and moderate differentiation among populations within drainages. Wright (1978) suggested that F_{st} values of 0.05 to 0.15 indicate moderate genetic differentiation, while values of 0.15 to 0.25 indicate great genetic differentiation. According to this scheme, differentiation between drainages is great within both *F. escambiae* (0.214) and within *F. nottii* (0.183). These numbers are not significantly different from each other and have broad overlap in 95% credible intervals (see figure 3.7). Differentiation between populations within the Pascagoula Drainage *(F. nottii)* is moderate (0.097), and between populations within the Blackwater drainage *(F. escambiae)* is moderate to great (0.156, with a large credible interval). These numbers imply that there is little migration of fish between drainages, which would be expected given the salinity tolerance of the fish. *F. nottii* only tolerates salinities up to 15 ppt (Crego and Peterson, 1997), whereas water of the Gulf of Mexico and even estuaries separating drainages has higher salinity (average 25 ppt in Choctawhatchee Bay estuaries, for example). The high θ^B estimates between populations within drainages may be an artifact due to small sample size, since they are based on only two drainages, each represented by only two populations. They may also be related to the fishes' natural history, however. Both *F. nottii* and *F. escambiae* prefer stream margins and backwaters, so adults may be unlikely to migrate through flowing areas of a river system. No information could be found

regarding larval behavior. More extensive sampling is needed to verify that within-drainage θ^B in these fish is indeed this high.

Error analysis

The minority of studies using AFLP data present an analysis of error rate of the method, but a sample of 13 that did present error data revealed that anywhere from 0 to 7.9% of scores differ between repeated runs of the same individual (Tohme *et al.,* 1996; Hartl and Seefelder, 1997; Hongtrakul *et al.,* 1997; Donaldson *et al.,* 1998; Busch *et al.,* 2000; Donaldson *et al.,* 2000; Bagley *et al.,* 2001; Congiu *et al.,* 2001; Gulao *et al.,* 2001; Miller *et al.,* 2002; Mock *et al.,* 2002; Kruse *et al.,* 2003; Whitehead *et al.,* 2003). Some studies also found that DNA of some individuals fails to work in the AFLP process (Palacios *et al.,* 1999; Busch *et al.,* 2000; Donaldson *et al.,* 2000; Bensch and Åkesson, 2005). The 6.1% level of non-repeatability found in this study probably had a small effect on the analyses performed on the AFLP data, except that it made recovery of species-diagnostic loci unlikely. Assuming an equal probability that any given score is incorrect, allele frequencies would be skewed in the direction of 50% on average (given actual genotypes at a locus, it is more likely that one of the more common score types will be observed incorrectly), increasing polymorphism and decreasing θ^B and distances between populations and groups. Simulation studies could be performed to quantify the effect such an error rate is likely to have.

Comparing population genetic structure between species

Population genetic parameters are compared for *F. nottii* and *F. escambiae* in table 3.3. *F. nottii* has significantly more genetic variability and perhaps less subdivision than *F. escambiae* (θ^B values are not significantly different, however). *F. nottii* clearly shows no genetic isolation by distance, whereas *F. escambiae* shows distinct isolation by distance. This pattern is even reflected in the mitochondrial sequence data, where *F. escambiae* sequences group by drainage, but *F. nottii*

Table 3.3. Comparison of population genetic parameters between the *F. nottii* and *F. escambiae.* Numbers significantly different are marked with an asterisk.

Parameter	*F. nottii*	*F. escambiae*
Percent polymorphic loci	68.9*	61.1*
Genetic diversity measured by heterozygosity	0.241*	0.209*
Pairwise distances between drainages and populations within drainages (in parentheses)	0.066 (0.050)	0.063 (0.030)
θ^B at the level of drainage within group and population within drainage (in parentheses)	0.183 (0.097)	0.214 (0.156)
Isolation by distance	r = 0.289 p = 0.119	r = 0.631 p = 0.051

sequences do not. Taken together these data strongly suggest there is higher movement of individuals between populations in *F. nottii* than in *F. escambiae* and implies that population genetic structure has likely changed in at least one of these species from the ancestor. It cannot be inferred whether the change happened during or subsequent to speciation, nor which group changed from the ancestral condition.

Morphological data (Wiley, 1977) and combined morphological and molecular data (Ghedotti and Grose, 1997) indicate that *F. lineolatus* is the closest relative to the sister pair *F. escambiae* and *F. nottii.* Assessment of the population genetic structure in this third species could provide evidence as to which member of the pair has diverged most from the ancestral condition. Population genetic data from another species pair formed in the same event could provide evidence bearing on the question of whether the speciation event itself brought about the change, or whether the change is simply do to characteristics unique to the species involved.

Chapter 4: Comparison of population genetics of *Fundulus* and *Ammocrypta* species pairs—Does a shared speciation event leave commonalities?

The goal of this chapter is to compare previously-derived population genetic parameters for two geminate pairs of species *(Ammocrypta beanii/A. bifascia* and *Fundulus nottii/F. escambiae)* to assess whether a shared speciation event resulted in common population genetic patterns between the pairs.

Introduction

Rationale

Can shared speciation events leave common population genetic patterns? Or are population genetic characteristics of geminate species pairs unchanged through speciation or shaped predominantly by characteristics of the individual taxa? A shared speciation event means the dividing of two or more ancestral species into geminate species pairs in the same time and place subsequent to the same cause of genetic division. It entails sharing a vicariant event and sharing a geography. The vicariant event is the actual splitting of ancestral populations into isolated groups, cutting off gene flow – often caused by a geological event such as mountain building or stream capture. In species with similar habitat requirements, in this case freshwater streams, a shared vicariant event is likely to result in similar rates of cessation of gene flow, similar completeness of isolation, and similar geographic progression of isolation over time (*i.e.,* In which part of the range do populations become isolated earliest, and in which direction does isolation progress?). These factors will be referred to as the dynamics of the isolation event. Species that share a

speciation event also share a geography simply by virtue of being located in the same region of the landscape. A shared geography, in species with similar habitat requirements, can mean similar impediments to migration and similar amount and availability of habitat. As explained in the introduction (see figure 1.2) comparison of replicate species pairs allows inferences to be drawn about the population genetic effects of the speciation event itself. Evidence that the speciation process itself has affected population genetic structure is seen when the member species of each pair differ in a population genetic parameter such that the species from each pair with the higher value of the parameter are both found on one geographic side of the vicariant line and the member of each pair with the lower value on the other (figure 4.1). Such

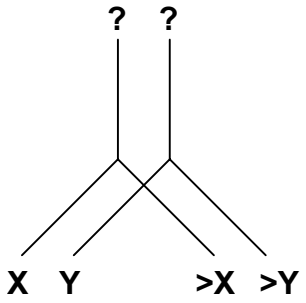


Figure 4.1. Evidence for "geography rules," or that the speciation event itself brought about change in the population genetic structure of the descendant species. Diagram similar to that shown in Chapter 1, figure 2. Shown are two phylogenies with ancestral species at the top and paired descendants below. The western descendant of the ancestor shown to the left has the value X for a population genetic parameter, while the eastern descendant has a value that is greater than X. Similarly, the western descendant of the ancestor shown to the right has the value Y for a population genetic parameter, while the eastern descendant has a value that is greater than Y. The question marks at the top indicate that the ancestral values of the parameters cannot be inferred. Given this scenario, however, it is most parsimonious to infer that the speciation process itself affected population structure in the descendant species (either caused a raising of the value in the east, or a lowering of it in the west).

a situation will be referred to as two pairs of sister species having a similar 'pattern of change' in the genetic parameter in question. In the null case, the species pairs could be expected to have either no changes in population genetic structure subsequent to speciation or changes idiosyncratic to species, dictated by natural history of the species involved or chance zoographic or demographic events. The alternative hypothesis is that "geography rules," or the shared isolation dynamics and geographical features of the landscape (speciation) have a defining role in shaping the population genetic characteristics of the geminate species pairs formed. Specific null and alternative hypotheses can be identified relative to this general framework.

Hypotheses

When different species pairs are formed by the same vicariant event, population genetic parameters—such as amount of genetic variation within species, genetic subdivision within each species, amount of genetic isolation by distance and divergence among populations within each species—might (1) be influenced by shared geography and history of divergence ("geography rules") or (2) might be left unchanged by the speciation event or more strongly influenced by species-specific idiosyncrasies (a two-part null case). If "geography rules," the species pairs would show a similar pattern of change in the feature in question, while the null hypothesis (not listed for each hypothesis) is that the two taxonomically distant species pairs, even if products of the same vicariant event, are not expected to show a similar pattern of change in the feature in question. Hypothesis 1 does not fit the pattern

described above as it deals with genetic distance between species rather than with an aspect of genetic structure.

Hypothesis 1: Two taxonomically distant species pairs will show similar levels of between-species genetic divergence if they formed in the same vicariant event.

Hypothesis 1 deals with the amount of genetic divergence between the two species in each pair. If rate of genetic divergence is shaped by the dynamics of the isolation event and the geography in which the organisms are found, two species pairs formed in the same vicariant event would show the same amount of genetic divergence between the species in each pair. If, on the other hand, rate of genetic divergence is more strongly influenced by characteristics unique to each species pair, rate of genetic divergence may not be the same among the species pairs, and different genetic distances may be found between the species in each pair. Characteristics that may be unique to each species and that could influence rate of genetic change include effective population size, mobility between populations, particular phylogenetic history, genetic makeup, generation time and selection acting on the species.

Hypothesis 2: The pattern of change in genetic distance between subgroups within two taxonomically distant species pairs will be similar if they formed the same vicariant event.

Hypothesis 3: The pattern of change in genetic subdivision within two taxonomically distant species pairs will be similar if they formed in the same vicariant event.

Hypothesis 4: The pattern of change in amount of genetic variability within two taxonomically distant species pairs will be similar if they formed in the same vicariant event.

Hypothesis 5 refers to genetic distances between pairs of populations, where one population of the pair is from each of a pair of sister species. Distance is calculated between pairs of populations located in close proximity to the vicariant dividing line and between pairs of populations located at as great a distance as possible from the dividing line. Then the ratio between the proximate population and distant population distances is calculated.

Hypothesis 5: Two taxonomically distant species pairs formed in the same vicariant event will be similar in such a ratio.

Hypothesis 5 addresses the dynamics of the isolation event that formed the two species pairs. To the degree that the isolation was or continues to be incomplete, hybridization may occur between the two species in each pair, resulting in increased genetic similarity in between-species comparisons of populations adjacent to the

vicariant dividing line. Populations at a distance from the vicariant dividing line would not be expected to show as much similarity because of the greater distance from any possible area of hybridization. As a measure of amount of hybridization and introgression, genetic distances can be compared between populations, chosen one from each species in a species pair, that are located (1) near the vicariant line versus (2) at a distance from the line. If the characteristics of the isolating event and the geography where the species are found determine the degree of hybridization and the degree to which introgressed genes travel throughout the ranges of the species, two distantly related species pairs sharing the same formative vicariant event would be expected to be similar in this ratio. If, on the other hand, these are predominantly determined by characteristics unique to the taxa involved, this ratio would not be expected to be the same between species pairs.

Hypothesis 6: Two taxonomically distant species pairs will show similar pattern of change in amount of within-species isolation by distance if they formed in the same vicariant event.

The species pairs included in this work

The two species pairs included in this work, *Ammocrypta beanii/A. bifascia* and *Fundulus nottii/F. escambiae,* comprise species of freshwater fishes living in the northern gulf coastal region of the United States (figure 4.2). Ranges of all four species are shown in figure 4.3. While the *Ammocrypta beanii/A. bifascia* species

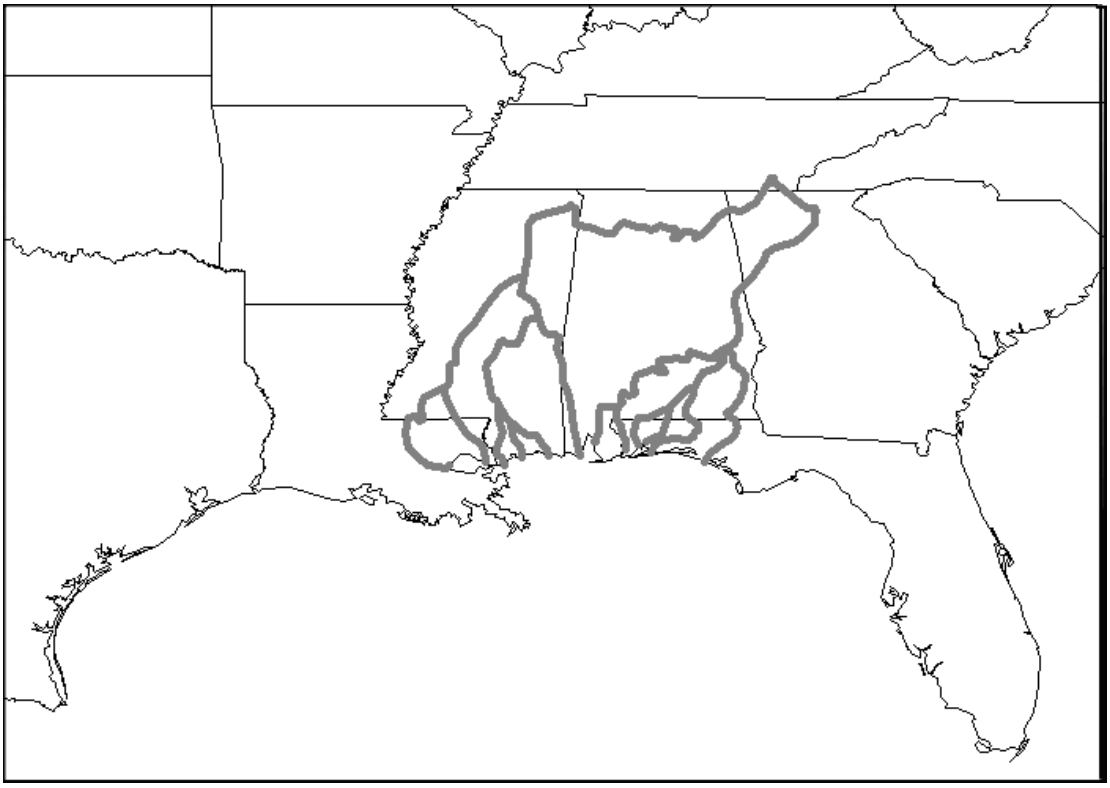


Figure 4.2. Approximate geographic location within the U.S. where ranges of included species lie. Boundaries of river drainages from which fish were sampled shown by grey lines.

pair presented unforeseen complications, with a distinct 3rd group present in the Mobile Bay Drainage (MOB) (figure 4.4 – see chapter 3 for details), AFLP (nuclear) data indicates that there was a divergence between *A. beanii* (including western and MOB groups) and *A. bifascia* with approximately the same geographic boundary as the divergence between *F. nottii* and *F. escambiae.* It is possible that these divergences occurred following the same geographic event, a possibility given support by the existence of multiple species pairs from diverse taxa with coincident species boundaries (Wiley and Mayden, 1985) as well as by geologic evidence of an event that could have brought about genetic isolation at approximately the correct

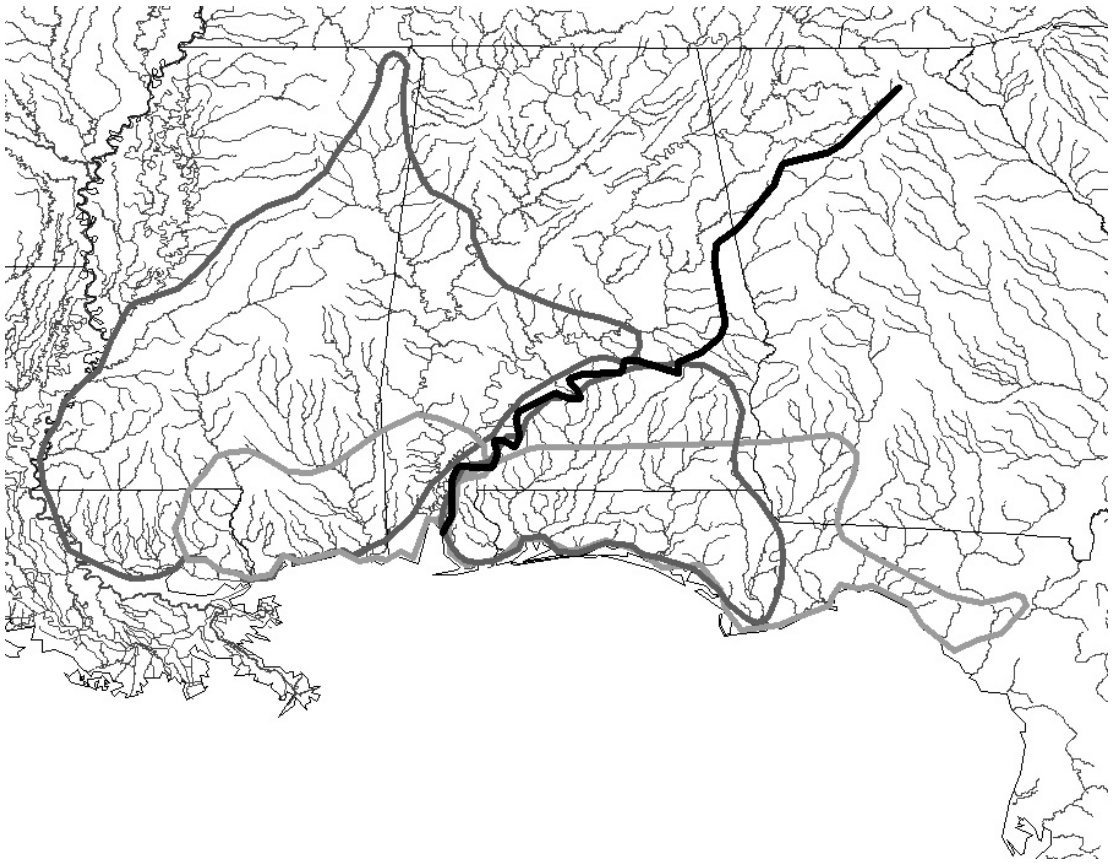


Figure 4.3. Ranges of the species pairs discussed in the study. *Ammocrypta beanii/A. bifascia* ranges shown in dark grey, *A. beanii* is to the west, *A. bifascia* to the east. *Fundulus nottii/F. escambiae* ranges shown in light grey, *F. nottii* to the west, *F. escambiae* to the east. The black line represents a possible vicariant dividing line, across which diverse other species pairs are also divided.

time (Price and Whetstone, 1977; Wiley, 1977). It is also possible that these divergences were independent, yet the species involved share the same geographical distribution.

These two species pairs are phylogenetically quite distant from each other. Both are members of the Acanthopterygii, higher spiny-rayed fishes, a huge group

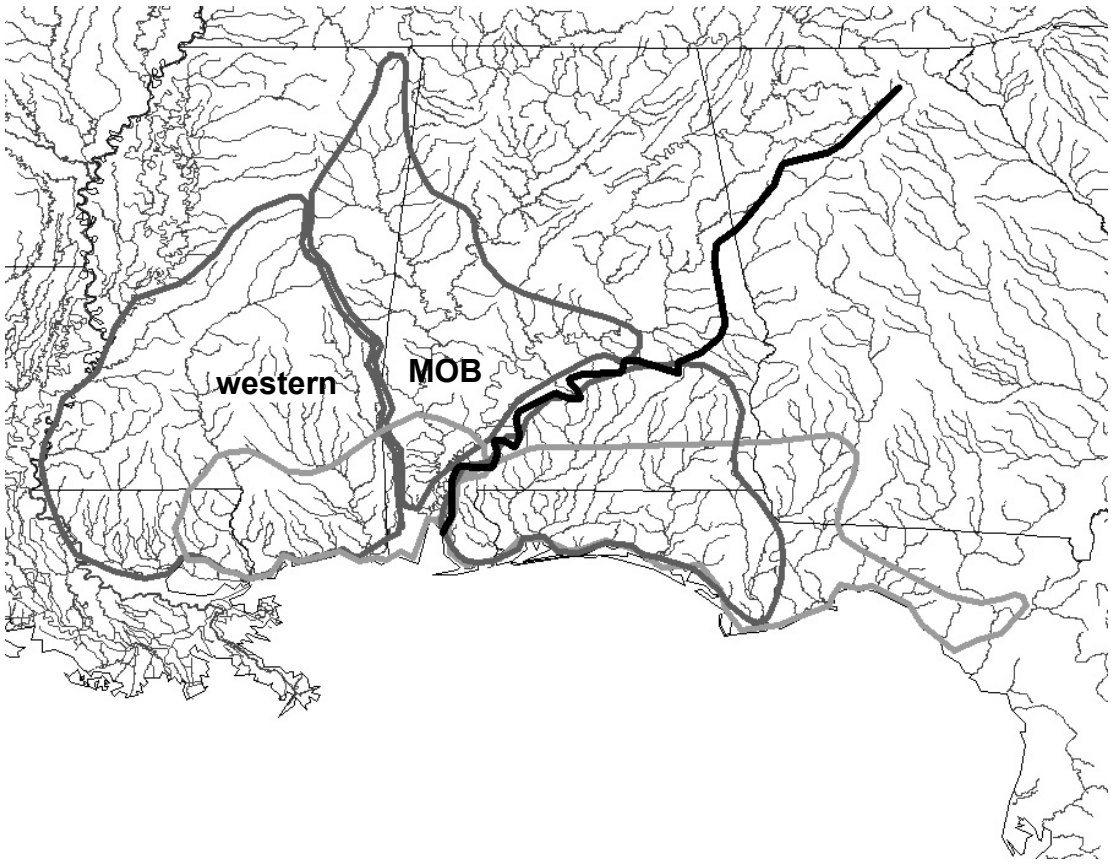


Figure 4.4. Ranges of the species pairs discussed in the study showing division of *Ammocrypta beanii* into western and MOB groups (labeled). *Ammocrypta beanii/A. bifascia* ranges shown in dark grey, *A. beanii* is to the west, *A. bifascia* to the east. *Fundulus nottii/F. escambiae* ranges shown in light grey, *F. nottii* to the west, *F. escambiae* to the east. The black line represents a possible vicariant dividing line, across which diverse other species pairs are also divided. AFLP (nuclear) data indicates that there was a divergence between *A. beanii* (including western and MOB groups) and *A. bifascia* with approximately the same geographic boundary as the divergence between *F. nottii* and *F. escambiae*.

containing approximately 13,500 species in 251 families (Helfman *et al.,* 1997). Family Fundulidae is located within the order Cyprinodontiformes, which is placed in the Atherinomorpha, a group located quite basally within the acanthopterygians. Family Percidae, on the other hand, which contains the genus *Ammocrypta,* is much higher in the acanthopterygian clade, within the order Perciformes. While the two

species within each pair, *F. nottii/F. escambiae* and *A. beanii/A. bifascia,* are quite similar natural history, there are distinct natural history differences between the two sister species pairs. More is known about the natural histories of *A. beanii* and *A. bifascia* than about *F. nottii* and *F. escambiae.* Even so, it is possible to point to distinct differences in biology between the two species pairs. All four species are common in the environments in which they occur, but the *Ammocrypta* species tend to be found at the bottom of the water column and buried in the sand in places of moderate to swift current. The *Fundulus* species, on the other hand, predominantly occupy backwaters, pools, and margins where there is little or no (sometimes moderate) current, and are found primarily near the surface of the water. The diet of the *Fundulus* is not well known, but they are surface-feeders while the *Ammocrypta* dart from hiding places on the bottom to capture prey. Little is known about the mating systems of these species, but in aquaria *F. nottii* males defend territories and mate singly with females (Baugh, 1981), although it is unknown whether a female might mate with more than one male serially. *F. nottii* eggs are laid on the bottom. No observation has been made of the breeding behavior of *A. bifascia* or *A. beanii*, but Johnston (1989) observed mating in a related fish (*A. pellucida*). This fish spawns on the bottom, with no nest made, and eggs are buried singly in the sand. Several males chase the female before spawning occurs, so it is unlikely that pairs form for the season, and seems that males and females are both promiscuous. All four species have a generation time of one year and may breed two or more times a year. Differences in life history and taxonomic divergence between the two species

pairs, coupled with their shared geography and likely shared vicariant event, makes comparison between them instructive when examining whether a shared speciation event or characteristics unique to species pairs involved plays a dominant role in shaping population genetic characteristics.

The vicariant event

Two different lines of evidence suggest that there was a general vicariant event across the eastern boundary of the Mobile Bay Drainage (MOB – figure 4.4), affecting diverse animal taxa. While many taxa in the area seem to have been unaffected by the event, Wiley and Mayden (1985) documented seven species pairs of fishes and three of snakes with one member of the pair geographically located in MOB and (usually) drainages to the west, and the other member of the species pair located in drainages to the east of MOB. Wiley and Hagen (1997) noted that two species pairs of fishes originally thought to contradict this pattern by Wiley and Mayden (1985) actually were distributed in accordance with the pattern, as well. Coincident species distributions of this kind are one line of evidence suggesting a vicariant event occurred across the eastern boundary of the MOB. The other line of evidence is the geological work of Price and Whetstone (1977), which was first connected to a possible vicariant event by Wiley (1977) in relation to the species pair *Fundulus nottii/F. escambiae*. They documented evidence for a Gulf-ward migration of flow of rivers to the east of the MOB (including the Escambia and Choctawhatchee rivers) that would have separated these river drainages from those further west by the late Pliocene or early Pleistocene, sometime during the last 5 million years. This

change in flow direction probably separated these drainages from those further west even during periods of sea-level lowering that have occurred in the last few million years. During these periods of sea-level lowering, river drainages would be elongated on the extended coast and drainages currently separated by salt water could connect, allowing migration of freshwater fish between what are now isolated drainages. This separation of the rivers could have genetically isolated freshwater fishes in the MOB and drainages to the west from river drainages to the east of MOB.

Methods

The population genetic parameters average heterozygosity, average pairwise genetic distance between populations, genetic isolation by distance, and θ^B (a Bayesian analog of F_{st}) were derived as discussed in the methods sections of Chapters 2 and 3. In addition, genetic distances between population pairs in close versus distant geographical proximity to the vicariant line were compared for this chapter. The method for this analysis is described in the next section. *Ammocrypta beanii* and *A. bifascia* samples are those described in Chapter 2. *Fundulus nottii* and *F. escambiae* samples are those described in Chapter 3.

Genetic distance between near and far pairings of populations

In this analysis four populations were selected from each species, two from near the boundary with the sister species and two distant from it. See figure 4.5 for maps showing the drainage boundaries and locations of the 'near' and 'distant' populations selected for each species. All populations chosen were represented by ten or more individuals, with an average of 15. When possible, 'near' populations

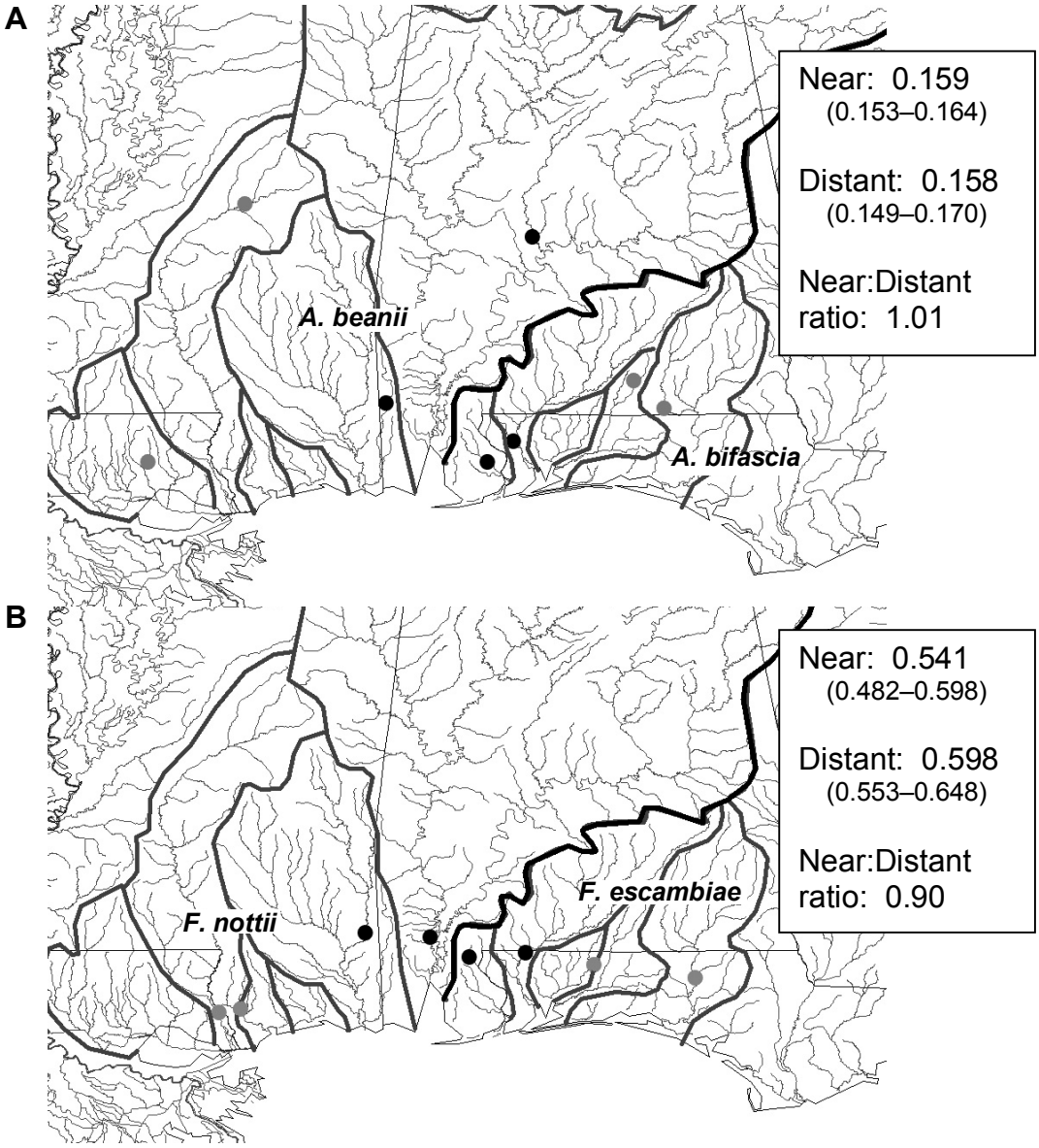


Figure 4.5. Populations used in comparison of genetic distance between cross-species pairs of geographically near versus distant populations. (A) shows populations chosen from *Ammocrypta beanii/A. bifascia;* (B) shows those from *Fundulus nottii/F. escambiae.* Black line represents the dividing line between ranges, grey lines show drainage boundaries. Black dots show populations used in 'near' pairs (two populations per species), while grey dots represent populations used in 'distant' pairs (again two per species). In the comparison, a 'near' pair comprised one black population from each species in a species pair, while a 'distant' pair comprised one grey population from each species in a species pair. Numbers to side are average and range of the four pairwise comparisons possible in each category for each species pair, and their ratio.

were chosen from drainages adjacent to drainages occupied by the sister species and 'distant' populations were selected from drainages not bordering drainages occupied by the sister. This was not possible in all cases. In this analysis, the population genetic software PopGene 1.32 (Yeh *et al.*, 1997) was used to derive Nei's (1978) unbiased genetic distance between pairs of populations based on AFLP data. Each population pair consisted of one population from each species in a sister species pair. Two sets of distances were derived for each species pair: pairwise distances between 'near' populations (represented by black dots on figure 4.5) and pairwise distances between 'distant' populations (represented by grey dots on figure 4.5). Since two 'near' and two 'distant' populations were chosen per species, four distances were derived in each category for each species pair (one for each possible pairing of the two populations in one species with the two populations in the sister species). Because these distances are not independent, confidence intervals on their averages could not be calculated. Ratio of the average 'near' distance to the average 'distant' distance was calculated for each species pair as a measure of whether populations near the species boundary show any more genetic similarity to the sister species than do populations distant from the species boundary.

Results

Genetic distance between species – Hypothesis 1

Nuclear as well as mitochondrial genetic distances were several times greater between *F. nottii* and *F. escambiae* than between *A. beanii* and *A. bifascia* (table 4.1). For comparison with sequence distances, AFLP distances are presented with all

Table 4.1. Genetic distances between species in each species pair.

	A. beanii/A. bifascia	*F. nottii/F. escambiae*
K2P distance between cytochrome *b* sequences	0.0117	0.0628
Nei's (1978) unbiased distance based on AFLP data	0.114	0.467

individuals within species treated as one group rather than clustered into populations.

Finding a valid comparison

Figure 4.6 shows Nei's (1978) unbiased genetic distances based on the AFLP data, with *F. nottii* divided into western and MOB regions for comparison with these groups in *A. beanii*. The values shown are averages of pairwise distances between drainages. The distance between the western and MOB groups in the case of *A. beanii* is 70% as large as that between the two species (0.117 versus 0.167 or 0.162 depending on the *A. beanii* group compared), and it is about twice as large as the distance between drainages within either the MOB or western group. This shows that there is considerable genetic division between the MOB and western *A. beanii* groups. In the case of *F. nottii,* however, the distance between the western and MOB groups is only about 15% of the distance between the two species and is not much larger than the distances between drainages in the western group, showing much less division between western and MOB regions in the case of *F. nottii.* Because there is a large subdivision between western and MOB populations in *A. beanii* and no corresponding division in *F. nottii,* the most appropriate population genetic

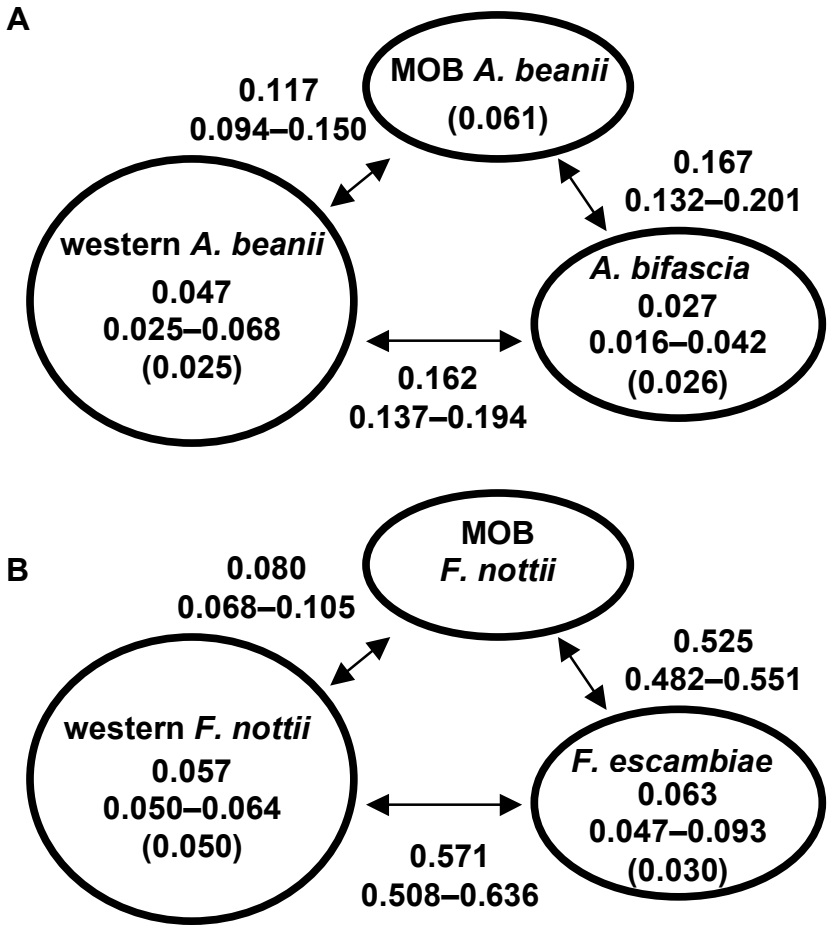


Figure 4.6. Nei's (1978) unbiased genetic distances among geographical groups calculated in PopGene 1.32 (Yeh *et al.,* 1997) based on AFLP data. (A) shows *Ammocrypta* groups, (B) shows *Fundulus* groups. Numbers in figure are average pairwise distances between drainage pairs selected one from each group (minimum and maximum also given). Numbers in parentheses are average pairwise distance between populations within drainages for the groups. Compared to the distance between species, distance between western and MOB groups is insignificant in the case of *Fundulus nottii,* but substantial in the case of *Ammocrypta beanii.*

comparison between species pairs involves western *A. beanii* or the MOB drainage (not both combined) to represent *A. beanii*. Once the schism in *A. beanii* between MOB and western drainages is recognized, and seen to be unlike the condition found

in the *Fundulus* species pair, the question arises as to how these groups compare aside from this difference. As discussed in Chapter 2, MOB *A. beanii* is sparsely sampled, and results for it are likely to be unreliable. Therefore, values relevant to the other hypothesis pairs will be shown for the groups western *A. beanii* and *A. bifascia* for comparison with the *F. nottii/F. escambiae* species pair.

Genetic distance between populations – Hypothesis 2

Figure 4.7 shows within group and between group genetic distances. Distance between the *Fundulus* species is more than three times that between western *A. beanii* and *A. bifascia.* Western *A. beanii* has higher average distance between drainages

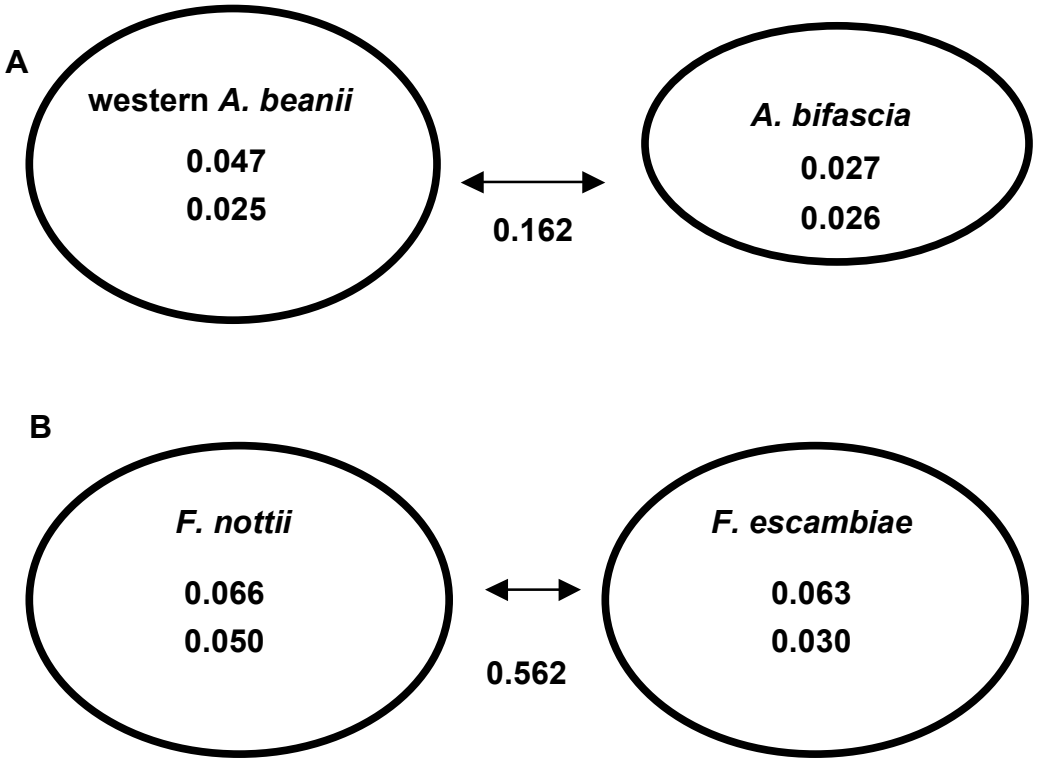


Figure 4.7. Nei's (1978) unbiased genetic distances between groups calculated in PopGene 1.32 (Yeh *et al.,* 1997). (A) shows *Ammocrypta* groups, (B) shows *Fundulus* species. Upper numbers are average pairwise distances among drainages within groups. Lower numbers within circles are average pairwise distance among populations within drainages.

than *A. bifascia,* while *F. nottii* and *F. escambiae* have very similar average distance between drainages. Distances between populations within drainages, on the other hand, are approximately the same between western *A. beanii* and *A. bifascia,* but are higher for *F. nottii* than for *F. escambiae.*

Population subdivision – Hypothesis 3

Figure 4.8 shows θ^B (the Bayesian F_{st} analog) values at different levels of analysis for western *A. beanii* and *A. bifascia* (A) and the *Fundulus* species (B). At

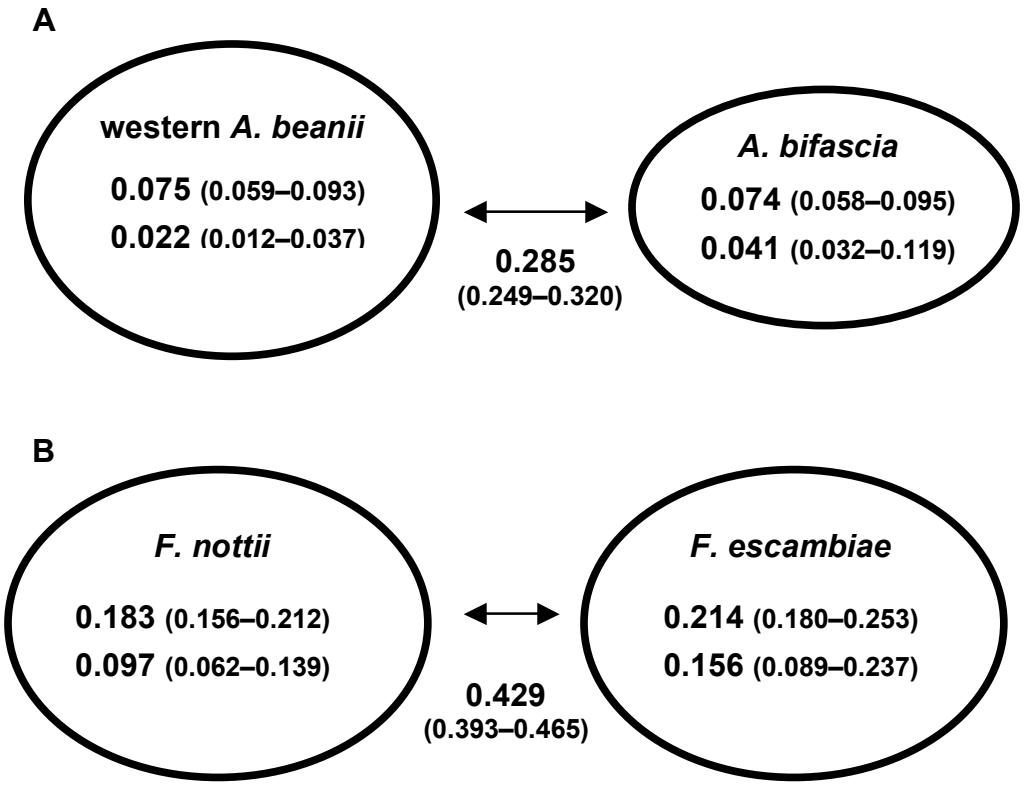


Figure 4.8. θ^B at different levels of analysis calculated in the program Hickory (Holsinger *et al.,* 2002). (A) *Ammocrypta* groups, (B) *Fundulus* species. θ^B at species level shown between groups. Within species, upper number is drainage within species and lower number is population within drainage. Numbers in parentheses are 95% credible intervals.

the highest level, subdivision is much larger between *F. nottii* and *F. escambiae* than between western *A. beanii* and *A. bifascia.* θ^B is also much larger within *F. nottii* and *F. escambiae* than within the *Ammocrypta* both the level of drainage within larger group and the level of population within drainage. Subdivision between drainages is the same in western *A. beanii* as it is in *A. bifascia,* and not significantly different between *F. nottii* and *F. escambiae* (95% credible intervals overlap broadly). At the population within drainage level, both eastern groups *(A. bifascia* and *F. escambiae)* have higher subdivision, although this is only significant in the *Ammocrypta* groups.

Genetic variation – Hypothesis 4

Figure 4.9 shows heterozygosity and polymorphism estimates for *Fundulus* and *Ammocrypta,* both total values for the pair and for each group within the pair. Both heterozygosity and polymorphism are higher for *Fundulus* than for *Ammocrypta* when total levels are measured. At the level of species, the western member in each pair has significantly higher polymorphism and heterozygosity than the eastern member. Comparing across species pairs, polymorphism does not differ between western species (western *A. beanii* and *F. nottii)* or between eastern species *(A. bifascia* and *F. escambiae),* but heterozygosity is higher in the *Fundulus* than in the *Ammocrypta.*

Genetic distance between proximate and distant pairs of populations – Hypothesis 5

Populations near the geographic boundary with the range of the sister species did not show any more genetic similarity with populations of the sister species than did populations far from the boundary in the *A. beanii/A. bifascia* species pair, but

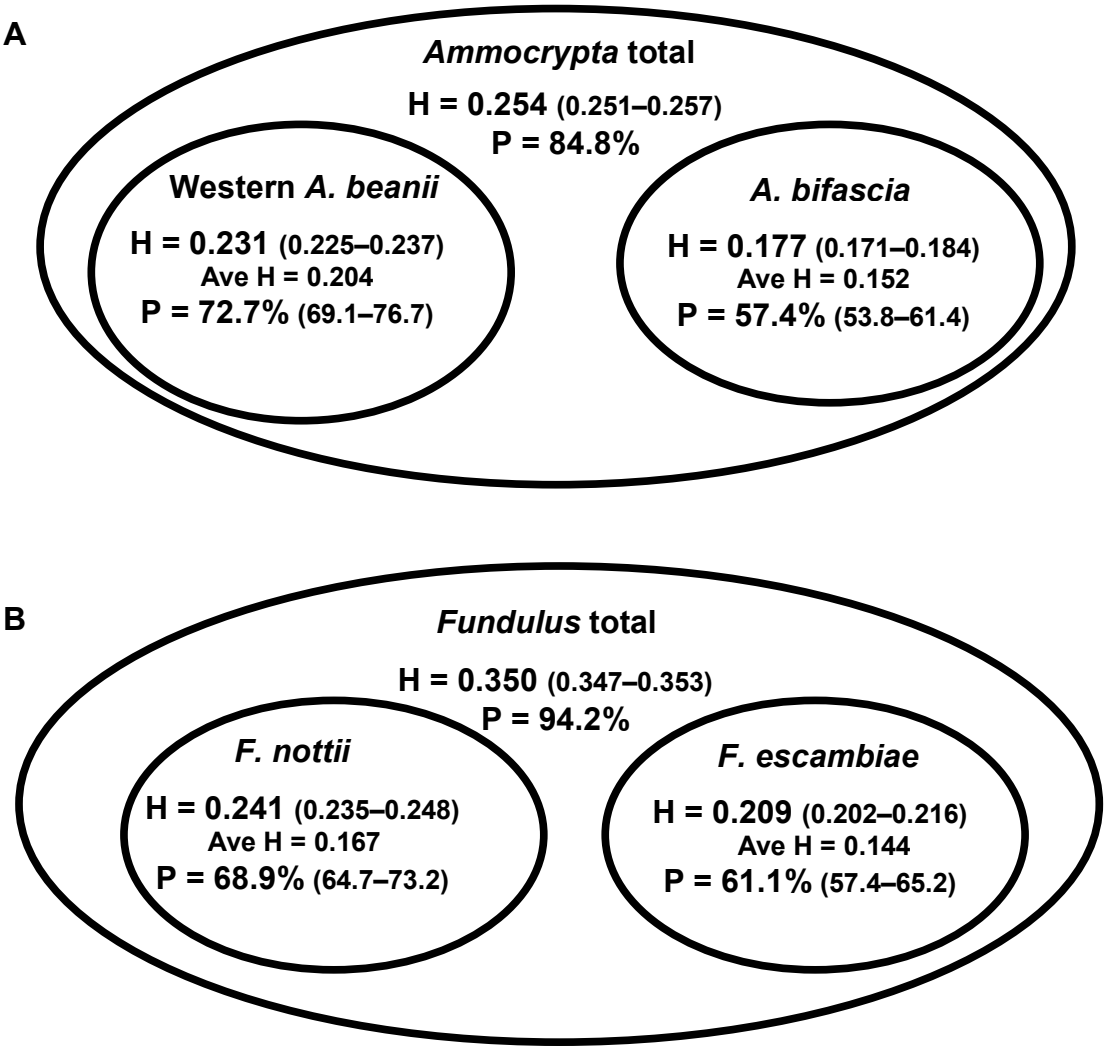


Figure 4.9. Heterozygosity and polymorphism. (A) *Ammocrypta*, (B) *Fundulus.* H is average heterozygosity calculated for the group as a whole and Ave H is the average of population-level average heterozygosities. P is percent polymorphic loci at the 95% level (most frequent allele frequency <95%). Heterozygosity estimates have 95% credible intervals shown and polymorphism estimates have 95% confidence interval shown.

there was a greater range in the genetic distances between 'distant' population pairs (figure 4.5). In case of the *F. nottii/F. escambiae* species pair, the average distance

between 'close' populations was only 90% as great as that between 'distant' populations.

Isolation by distance – Hypothesis 6

Comparison of isolation by distance between the groups is shown in figure 4.10. Western *A. beanii* shows a higher correlation between geographical and genetic distance than *A. bifascia,* but the correlation in western *A. beanii* is only bordering on significance ($p = 0.06$). There is no isolation by distance in *F. nottii* ($p = 0.119$), while there is a correlation of 0.63 between genetic and geographic distance in *F. escambiae* ($p = 0.051$).

Discussion

Genetic distance between species – Hypothesis 1

That genetic distance between *F. nottii* and *F. escambiae* is several times greater than that between *A. beanii* and *A. bifascia* indicates one or a combination of several possibilities: (1) the *Fundulus* species pair experienced genetic isolation earlier than the *Ammocrypta* pair, (2) the *Fundulus* species have been diverging genetically at a faster rate than the *Ammocrypta* pair and (3) the *Fundulus* already had substantial genetic differentiation between geographic regions before the vicariant event occurred. Relative to possibility (1), the geographical evidence of Price and Whetstone (1977) and the biogeographical evidence of Wiley and Mayden (1985) suggest that the species pairs formed in the same vicariant event, although this is not a certainty. If they did form in the same event, AFLP data indicate that the rate of change in the *Fundulus* nuclear genome would have had to be 3.4 times faster than

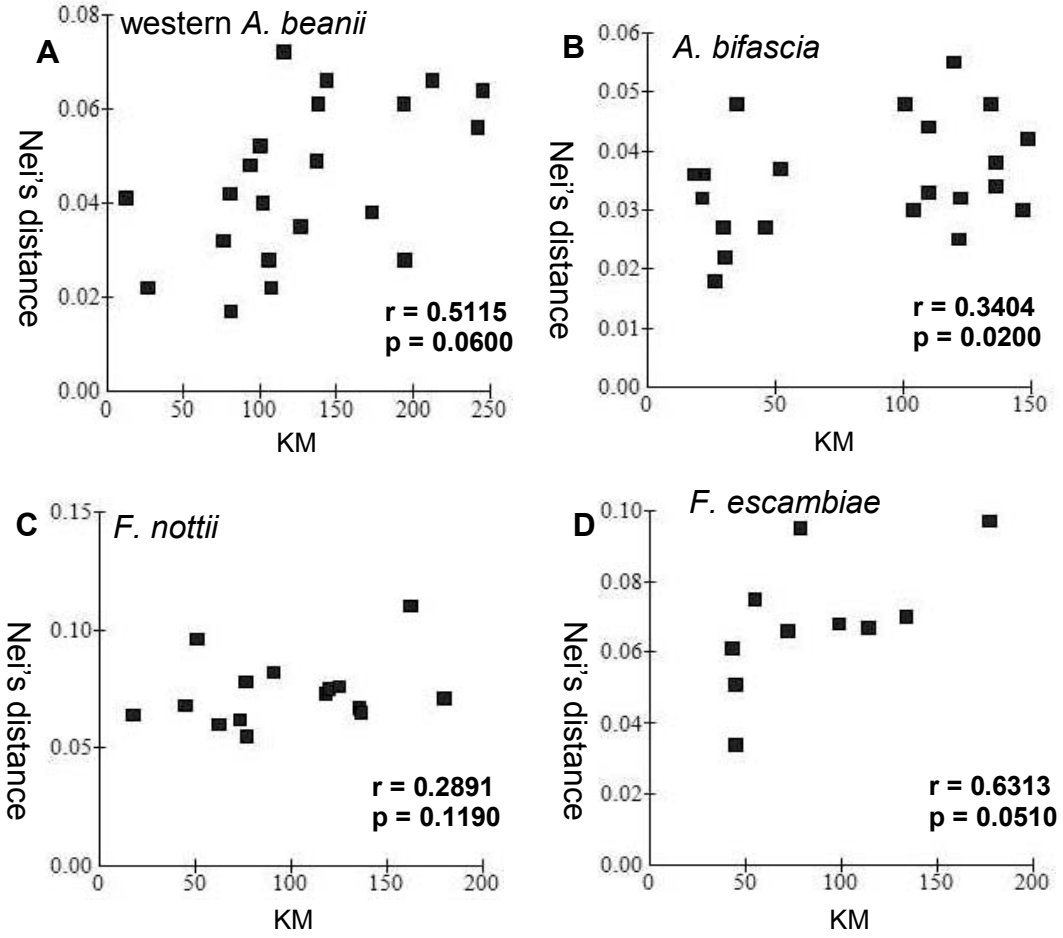


Figure 4.10. Plot of Nei's (1978) genetic distance versus geographical distance between populations as assessed in a Mantel test (A) for western *Ammocrypta beanii;* (B) for *A. bifascia;* (C) for *Fundulus escambiae*; (D) for *F. nottii.* The Mantel test was conducted in TFPGA (Miller 1997).

the rate of change in the *Ammocrypta* nuclear genome (possibility (2) — table 4.1). No literature can be found about overall mutation rates in nuclear genomes of poikilothermic vertebrates, although differences in rate of molecular evolution among lineages are well documented (*e.g.,* Li and Tanimura, 1987; Gu and Li, 1992; Martin and Palumbi, 1993; Yang and Nielson, 1998; Bromham, 2002). According to Yang and Nielson's (1998) data for 48 different genes, the ratio of rodent to human

synonymous substitutions is 2.8, although this may relate to different generation times between humans and rodents. Using a relative rates test with allozyme data, Osinov and Lebedev (2000) found genetic distance from *Oncorhychus* (Pacific salmon) to an outgroup was 1.5 to 1.75 times greater than from *Hucho, Salvelinus,* and *Parahucho* (other genera in the same family) to the outgroup. Given that about half the distance (from the outgroup to the ancestor) was constant in each comparison, this suggests that the rate of evolution of *Oncorhychus* for these loci was about three times that of the other genera. Although data on relative rates of overall nuclear genome evolution in vertebrates is lacking, these cases involving specific genes indicate that a 3-fold difference in rate between two groups of fishes in different orders, such as *Fundulus* and *Ammocrypta,* is conceivable. Possibility (3) also seems plausible; the *Fundulus* taxa have much higher genetic subdivision than to the *Ammocrypta.*

If these two species pairs did form in the same vicariant event (possibilities (2) and (3)), shared isolation history and geography did not play a dominant role in determining amount of genetic distance between the species pairs and the null hypothesis is not rejected. New research questions arise in investigating what species-specific characteristics play a role in determining rate of genetic change. Biological characteristics hypothesized to shape rates of molecular evolution include body size (Martin and Palumbi, 1993), generation time (Wu and Li, 1985; Gaut *et al.,* 1992; Li *et al.,* 1996), body temperature (Martin and Palumbi, 1993), efficiency of DNA repair machinery, effective population size (Ayala, 2000), and even clutch size

(Bromham, 2002). *Ammocrypta* and *Fundulus* body sizes are similar, each group has a one-year generation time, they have the same body temperature, and there is no *a priori* reason to suspect a difference in DNA repair machinery between the two lineages. *A. bifascia* produce 19–85 eggs per clutch (Heins, 1985), *A. beanii* 27–120 (Heins and Rooks, 1984). No information is available for *F. nottii* or *F. escambiae,* but the related *F. dispar* produces 7–30 eggs per clutch (Taylor and Burr, 1996). According to the only study relating rate of molecular evolution to clutch size, smaller clutch sized in *Fundulus* should give them a *lower* rate of evolution (Bromham, 2002). Effective population size, then, is left as the explanation most likely to account for the faster rate of evolution of the *Fundulus* species. Smaller effective population size would tend to increase the rate at which neutral mutations become fixed in the genome (Ayala, 2000). Ross and Baker (1983) found a greater abundance of *A. beanii* than *F. nottii* in the Perdido Drainage, Heins and Rooks (1984) noted that *A. beanii* was the second most abundant species caught in the Biloxi Drainage (after *Hybopsis longirostris),* and Heins (1985) found *A. bifascia* to be the 3rd most abundant species in the Blackwater River Drainage (after *Hybopsis longirostris* and *Notropis versatus).* These observations may indicate that *A. beanii* and *A. bifascia* are generally more common than *F. nottii* and *F. escambiae,* although the habits of the *Fundulus* species would make them less common in specific areas of stream where the *Ammocrypta* species are found, and E.O. Wiley (personal communication) has observed both *Fundulus* species to be present in great numbers. The *Fundulus* tend to occupy still backwaters and pools while the *Ammocrypta* are

found in riffles and areas of moderate current. Factors other than numbers of individuals may be decreasing effective population size of the *Fundulus* species relative to the *Ammocrypta,* however. *Fundulus* don't venture into areas with current, a habit that may fragment their populations more than those of *Ammocrypta.* Breeding behavior also may have an effect. Based on observations of *A. pelucida,* in *Ammocrypta* both females and males are promiscuous, often with several males chasing one female at the same time (Johnston, 1989) but in *Fundulus,* the dominant male may primarily be the one that mates with females (Baugh, 1981). Thus the number of males contributing genes to the next generation may be reduced in *Fundulus* compared to *Ammocrypta.* Future studies are needed to address whether *Fundulus* do migrate less than *Ammocrypta,* and more observations of mating behavior are needed to assess whether there is greater variation in male reproductive success in *Fundulus* than in *Ammocrypta.* Greater population subdivision in the *Fundulus* species could also contribute to a greater divergence between species, as the populations originally separated from each other during the vicariant event would have been more differentiated. Although geographic and biogeographic data provide indirect evidence that the *Fundulus* and *Ammocrypta* species pairs formed in the same vicariant event, it cannot be ruled out that *F. nottii* and *F. escambiae* have greater genetic divergence than *A. beanii* and *A. bifascia* simply because they diverged earlier.

Genetic distance between subgroups – Hypothesis 2

In terms of genetic distance between subgroups, the greatest difference between the two species pairs is the great amount of genetic division between the MOB and drainages to the west in the western *Ammocrypta* species, *A. beanii,* with no corresponding division in the western *Fundulus* species, *F. nottii* (figure 4.6). Why is there a divergence in *A. beanii* between the MOB and drainages to the west while such a pattern does not appear in *F. nottii?* This question reflects a larger biogeographic question among freshwater fishes in the southeastern United States: Why do some taxa and not others divide across apparent vicariant dividing lines? One example of this is the eastern boundary of the MOB, across which many taxa have become divided into species pairs including the species pairs analyzed in this work (Wiley and Mayden, 1985). Yet there are species with continuous distribution across the eastern boundary of the MOB. Future work should examine the characteristics of taxa that do versus those that do not divide across this line to see if there are any shared traits in either group that might explain the difference. Such work might also reveal possible reasons that *A. beanii* diverges across the western boundary of the MOB and *F. nottii* does not.

When western *A. beanii/A. bifascia* is compared to *F. nottii/F. escambiae,* the pattern of genetic distance between drainages is different; average pairwise genetic distance between drainages is approximately the same in *F. nottii* and *F. escambiae,* but different between the *Ammocrypta* groups. The null hypothesis is not rejected; genetic distance between drainages either doesn't change during the speciation

process, or is primarily influenced by factors idiosyncratic to group. There is no support for the alternative hypothesis that the speciation event, itself, leaves a recognizable pattern in the species formed.

Population subdivision – Hypothesis 3

As shown in figure 4.9, there is no difference in θ^B at the level of drainage within species between western *A. beanii* and *A. bifascia* or between *F. escambiae* and *F. nottii*. At the level of population within drainage, a pattern is found that would support the alternative hypothesis: both eastern groups have values elevated over their western counterparts. This is only significant in *Ammocrypta*, however, and even there based on samples from only one or two drainages per species. Evidence at the level of population within drainage only weakly supports the alternative hypothesis, and is controverted by the evidence at the level of drainage within larger group. The null hypothesis should not be rejected. On the whole, these data suggest that population subdivision is unchanged through a speciation event, or shaped more by factors idiosyncratic to the species formed than by the dynamics of the speciation event itself.

At all levels of analysis, θ^B is much larger in the *F. nottii/F. escambiae* species pair than in the *Ammocrypta* groups. The factors discussed for hypothesis 1 that may decrease effective population size and increase genetic isolation in the *Fundulus,* would also produce this effect. It is reasonable to infer that the ancestral *Fundulus* species had higher genetic subdivision than the ancestral *Ammocrypta* species.

Genetic variation – Hypothesis 4

Of all the population genetic parameters estimated in this work, heterozygosity is the only one that clearly supports the alternative hypothesis that a speciation event may leave an identifiable population genetic pattern in geminate species formed. As shown in figure 4.9, both eastern groups *(A. bifascia* and *F. escambiae)* have lower heterozygosity and polymorphism than their western counterparts.

Total amount of genetic variation measured both as heterozygosity and percent polymorphic loci is higher in the *Fundulus* species pair than in the *Ammocrypta.* Once again, this could be caused by lower migration and lower effective population size in the *Fundulus.* These factors, while tending to decrease heterozygosity at the population level, cause populations to diverge from each other so that heterozygosity estimates increase when populations are combined into larger groups. Figure 4.9 shows that average population heterozygosity is lower in the *Fundulus* than in the *Ammocrypta.*

Genetic distance between pairs of populations – Hypothesis 5

No difference in the genetic distance between geographically distant versus close populations was found for the cross-species comparisons in the *A. beanii/A. bifascia* species pair, while geographically close populations were about 10% less genetically distant from each other in cross species comparisons in the *F. nottii/F. escambiae* species pair (figure 4.5). The simplest explanation for this is that there has been some fairly recent hybridization between *F. nottii* and *F. escambiae,* and none

(or much less) between *A. beanii* and *A. bifascia.* Another possible explanation is that introgressed genes spread more rapidly through the range of *Ammocrypta,* so that populations near the boundary between the two species are not perceptibly more similar to the sister species than populations at a distance from it. One method to distinguish between these two possibilities is to sequence several genes for multiple representatives of multiple populations within each species and look for individual sequences that seem to be found in the wrong species (introgressed due to hybridization). Then it could be seen directly whether only *Fundulus* shows evidence of hybridization or whether *Ammocrypta* shows movement of introgressed genes throughout the range. In either case, however, the null hypothesis is not rejected. There is no evidence that level of isolation by distance is directly shaped by the vicariant event or the geographic region.

Only two near and two distant populations were sampled for this analysis, however, and the geographic relationships between the populations were not the same for each species pair (figure 4.5). It is also possible that the different ratios calculated for the different species pairs were due to sampling error or differences in sampling regimes. Sampling of further populations could address this question.

Isolation by distance – Hypothesis 6

The correlation coefficient between genetic and geographic distance is larger in the western of the *Ammocrypta* groups and in the eastern of the *Fundulus* groups, although the significance is only 0.06 in the eastern *Ammocrypta* group. The null hypothesis is not rejected. Isolation by distance appears not to be directly influenced

by the vicariant event or the geographic region; rather by factors idiosyncratic to each species.

Conclusion

Figure 4.11 summarizes the results of the current study. Of the five population genetic parameters estimated, only one, heterozygosity, showed the pattern identified in figure 4.1 as supporting the hypothesis that a shared vicariant event and geography (speciation event) would leave a recognizable population genetic pattern in the species formed. Only two species pairs were analyzed here, one of which was distinctly split into two groups, but the evidence suggests that at least most population genetic parameters are shaped by factors idiosyncratic to species (such as zoography and historical demographic events). Speciation events may shape the pattern of heterozygosity in species pairs formed, although finding a looked-for pattern in one of five parameters could be due to chance. Population genetic analysis of further species pairs is necessary to assess whether this pattern is indeed caused by the speciation process.

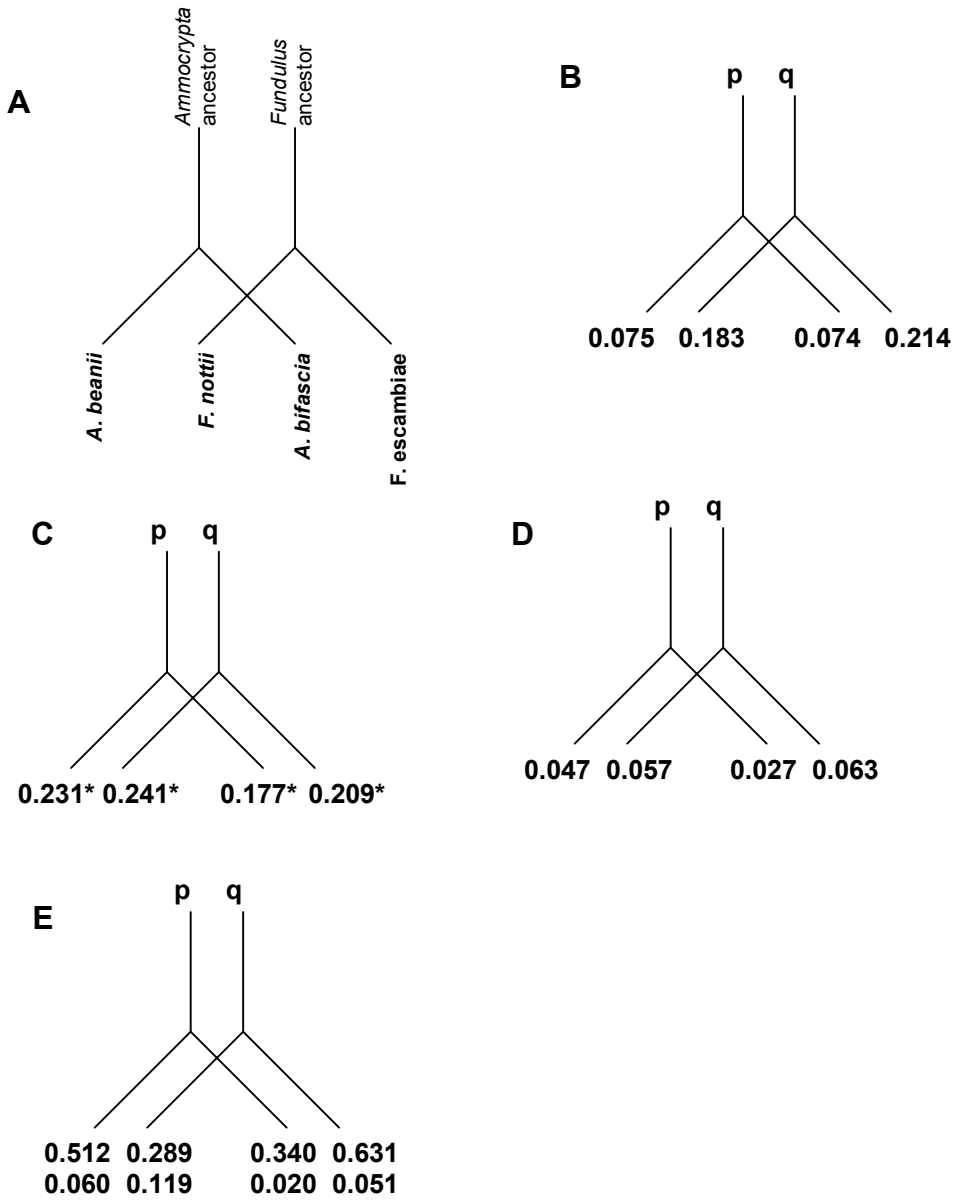


Figure 4.11. Summary of results of current study. *Ammocrypta* values shown on left, *Fundulus* on right. (A) species locations in the diagram; (B) θ^B (a Bayesian analog of F_{st}); (C) heterozygosity; (D) genetic distance between drainages (significance not known); (E) isolation by distance (upper number is correlation coefficient, lower number is p value). Starred values are significantly different between species in a species pair. Only heterozygosity shows the pattern supporting an effect of the speciation event itself on the species formed. Heterozygosity and θ^B were calculated in the program Hickory (Holsinger *et al.,* 2002), genetic distance was calculated in the program PopGene 1.32 (Yeh, *et al.,* 1997).

Chapter 5: Conclusion

This study asked whether observed population genetic structure of daughter species shows a pattern indicative of change caused by the speciation process itself or whether there is either no change or change idiosyncratic to species and presented a new method of analysis to address this question. Four species of fresh water fishes composing two geminate species pairs were examined: *Ammocrypta beanii/A. bifascia* (two species of sand darter) and *Fundulus escambiae/F. nottii* (two species of starhead topminnow). The pairs are thought to have formed in the same vicariant event in the southwestern United States. Population genetic parameters had never been analyzed for any of these species previously. This study incorporated both nuclear AFLP data and mitochondrial cytochrome *b* sequence data.

A large genetic division was found within the western *Ammocrypta* species *(A. beanii)* between populations within the MOB drainages and populations in drainages to the west of the MOB. Nuclear data grouped the MOB populations with the western *A. beanii* populations, while mitochondrial data grouped them with the sister species, *A. bifascia.* The nuclear data is considered more reliable, but the question of whether the MOB populations should be considered a new species is unresolved. No such genetic division was found in the western *Fundulus* species *(F. nottii).* The MOB populations of *A. beanii* were removed from comparative analysis between the species pairs to remove the effects of the genetic schism within *A. beanii.*

Distinct differences were found between the population structures of the two species pairs, with genetic distances and subdivisions generally much greater in the *F.*

escambiae/F. nottii species pair than in *A. beanii/A. bifascia*. The genetic distance between the *Fundulus* species was three times as large as that between the *Ammocrypta* species, both for the nuclear and for the mitochondrial data. Figure 5.1 summarizes population genetic statistics estimated for these four species. Population genetic estimates as well as phylogenetic relationships are shown on the diagram to facilitate comparison between species pairs. The greater amounts of subdivision and genetic distance between populations in the *Fundulus* species compared to the *Ammocrypta* species may relate to natural history of the species pairs. Topminnow affinity for backwaters may cause them to migrate less than sand darters, and their mating system likely decreases their effective population size more than that of sand darters. Because both species within the *Ammocrypta* pair were generally similar in population subdivision, it is parsimonious to infer that the ancestral *Ammocrypta* species had structure similar to that currently seen in its descendants. *A. beanii* had higher genetic variation than *A. bifascia*, however, making it clear that at least one of these species has changed from the ancestral condition in this characteristic. The ancestral condition cannot be determined, however, without analysis of an outgroup. Isolation by distance (and perhaps genetic subdivision) is greater in *F. escambiae* than in *F. nottii,* while genetic variability is lower. Again, it is clear that at least one of these species has changed from the ancestral condition. Again, an outgroup is needed to determine the ancestral state. No species-diagnostic loci were found for either species pair, and it was concluded that the AFLP procedure is inappropriate for looking for absolutely species-diagnostic loci because of an approximately 5% error

rate. Based on nearly diagnostic loci, there was a very high probability of correctly assigning an unknown specimen to its species of origin.

Of the four population genetic parameters estimated, only one, heterozygosity, showed the pattern identified in figure 4.1 as supporting the hypothesis that a shared vicariant event and geography (speciation event) would leave a recognizable population genetic pattern in the species formed. The speciation process appears to have had little or no effect on population subdivision and average genetic distance between populations in these two species pairs. Isolation by distance appears to have changed in ways idiosyncratic to individual species. Finding a looked-for pattern in one of four parameters examined could be due to chance and is not considered strong evidence of a link between speciation and microevolution. Only two species pairs were analyzed here, one of which was distinctly split into two groups. The results of this study suggest that microevolutionary parameters (allele frequencies and their distributions) are not directly shaped my macroevolutionary dynamics (the processes involved in speciation) in a predictable way, although more species pairs need to be examined before this result can be stated with confidence. This work presents a novel method for the study of the population genetic consequences of speciation.

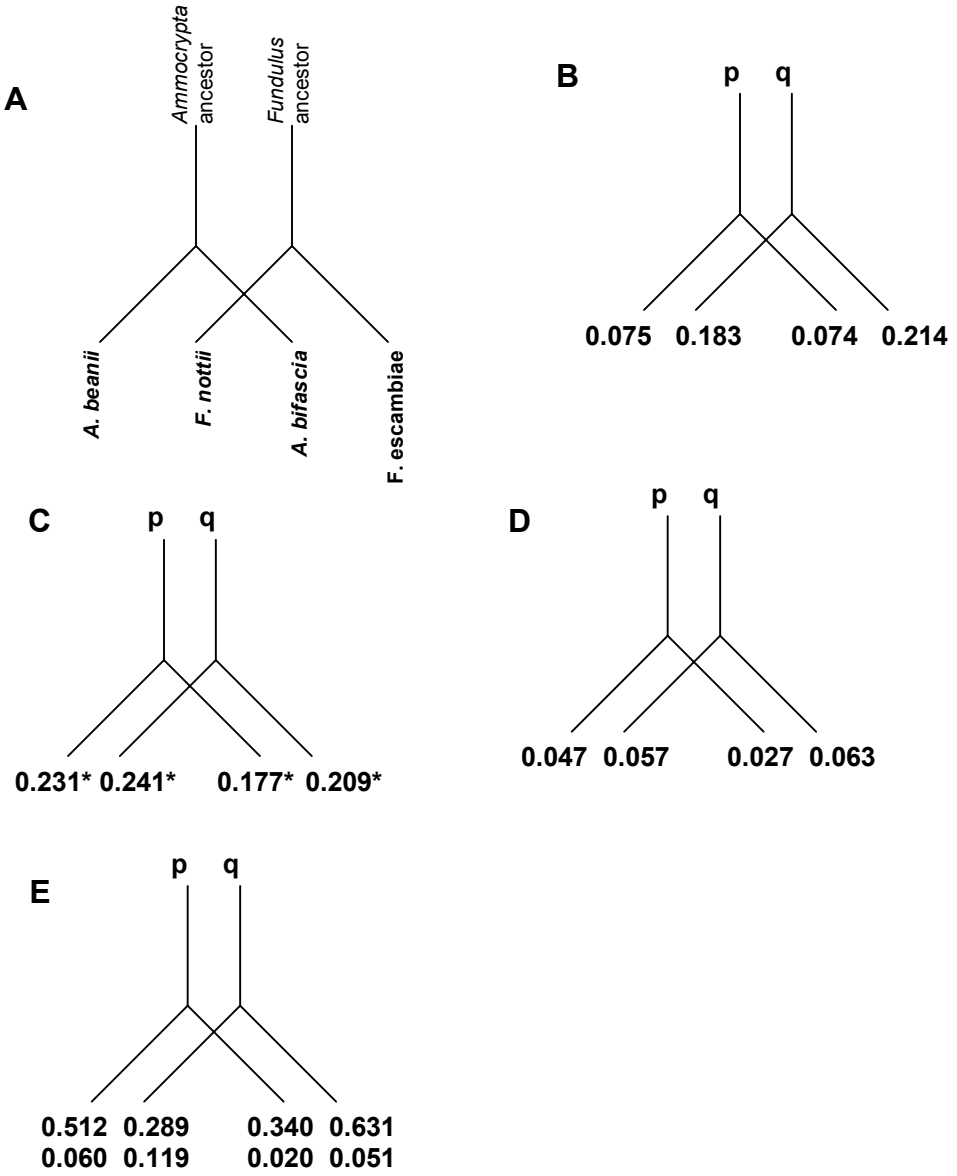


Figure 5.1. Summary of results of current study. *Ammocrypta* values shown on left, *Fundulus* on right. (A) species locations in the diagram; (B) θ^B (a Bayesian analog of F_{st}); (C) heterozygosity; (D) genetic distance between drainages (significance not known); (E) isolation by distance (upper number is correlation coefficient, lower number is p value). Starred values are significantly different between species in a species pair. Only heterozygosity shows the pattern supporting an effect of the speciation event itself on the species formed. Heterozygosity and θ^B were calculated in the program Hickory (Holsinger *et al.,* 2002), genetic distance was calculated in the program PopGene 1.32 (Yeh, *et al.,* 1997).

Appendices

Appendix 1. Samples used and sample success for *Ammocrypta* samples

Species	Drain.	Field #	KU Tissue #'s *	Proportion that worked
A. beanii	LPO	EOW 1988-08	884-1 to 884-10	10 of 10 (100%)
A. beanii	PRL	EOW 1989-09	832-1 to 832-3, 832-4*, 832-5 to 832-15	14 of 15 (93%)
A. beanii	BIL	EOW 1989-36	879-1*, 879-2 to 879-19	18 of 19 (95%)
A. beanii	PAS	EOW 1987-11	887-1 to 887-15	15 of 15 (100%)
A. beanii	PAS	EOW 1988-10	829-1 to 829-8, 829-10, 829-12 to 829-17	15 of 15 (100%)
A. beanii	PAS	EOW 1989-33	399-1 to 399-14, 399-27	15 of 15 (100%)
A. beanii	PAS	EOW 1988-09	828-1 to 828-8, 828-10, 828-11, 828-12*, 828-13, 885-1 to 885-3	14 of 15 (93%)
A. beanii	MOB (ALA)	EOW 1989-15	814-1 to 814-15	15 of 15 (100%)
A. beanii	MOB (TOM)	EOW 1989-16	604-1 to 604-5, 604-7	6 of 6 (100%)
A. bifascia	PER	EOW 1989-27	808-1 to 808-21	21 of 21 (100%)
A. bifascia	PER	EOW 1988-15	606 to 610, 889*	5 of 6 (83%)
A. bifascia	ESC	EOW 1989-25‡	877-1, 877-2*, 877-3 to 877-13, 616, 617, 618*, 619	15 of 17 (88%)
A. bifascia	ESC	EOW 1988-14	811-1 to 811-6, 989-1 to 898-10, 898-11*	16 of 17 (94%)
A. bifascia	BLA	EOW 1988-13	809-1 to 809-8, 809-11 to 809-13	11 of 11 (100%)
A. bifascia	YEL	EOW 1988-19	611*, 612 to 615, 812, 876, 886-1 to 886-9	15 of 16 (94%)
A. bifascia	CHO	EOW 1988-21	893-1 to 893-4, 893-5*, 893-6 to 893-11, 893-12*, 893-13, 893-14*, 893-15 to 893-19	17 of 19 (89%)
A. bifascia	CHO	EOW 1988-23	872-1 to 872-14	14 of 14 (100%)

* Sample didn't work

Appendix 2. *Ammocrypta s*pecimens used in phylogenetic analysis of cytochrome *b* sequence. Tissue number is given, followed by number of specimens used with that tissue number, followed by unique tissue identifications, if applicable. GenBank accession numbers for sequences are given in square brackets. GenBank accession number only is given for outgroup sequences.

MOB '*Ammocrypta beanii*'
KU T814, 3, (T814-7 [DQ512382], T814-8 [DQ512383], T814-9 [DQ512384]), voucher KU 24381, Alabama R.
KU T604, 2, (T604-1 [DQ512374], T604-2 [DQ 512375]), voucher KU 24382, Tombigbee R.

Ammocrypta beanii
KU T884, 3, (T884-1 [DQ512382], T884-3 [DQ512383], T884-5 [DQ512384]), voucher KU 29834, Tangipaho R., Lake Pontchartrain Drainage
KU T879, 3, (T879-8 [DQ512379], T879-16 [DQ512380], T879-17 [DQ512381]), voucher KU 29851, Big Biloxi R., Biloxi Bay Drainage
KU T829, 1, (T829-17 [DQ512371]), voucher KU 24384, Pascagoula R., Pascagoula Drainage
KU T399, 2, (T399-8 [DQ512372], T399-10 [DQ512373]), voucher KU 22898, Escatawpa R., Pascagoula Drainage

Ammocrypta bifascia
KU T898, 1, (T898-7 [DQ512388]), voucher KU 22146, Escambia R., Escambia Drainage
INHS 38107 [AF183940], 1, Burnt Corn Cr., Escambia Drainage
KU T809, 1, (T809-12 [DQ512387]), voucher KU 24862, Blackwater R., Blackwater Drainage
KU T610 [DQ512385], 1, voucher KU 24388, Perdido R., Perdido Drainage
KU T808, 1, (T808-1 [DQ512386]), voucher KU 24385, Perdido R., Perdido Drainage

Ammocrypta Meridiana—[AF183942]
Ammocrypta clara—[AF183941]
Ammocrypta pellucida—[AF183943].

Appendix 3. Samples used and sample success for *Fundulus* samples

Species	Drain.	Field #	KU Tissue #'s *, †	Proportion that worked
F. nottii	PRL	EOW 1989-39	524-1 to 524-4 (All *), 6721 to 6726, 6727*, 6728 to 6731	14 of 19 (74%)
F. nottii	BST	EOW 1989-37	6705, 6706*, 6707*, 6708 to 6711, 6712*, 6713*, 6714 to 6720	12 of 16 (75%)
F. nottii	BIL	EOW 1989-35	6689 to 6703, 6704*	15 of 16 (94%)
F. nottii	PAS	EOW 1989-34	542-1 to 542-5 (All *), 629-1*, 629-2*, 6671 to 6675, 6676*, 6677 to 6688	17 of 25 (68%)
F. nottii	PAS	EOW 1987-12	6637*, 6638 to 6652	15 of 16 (94%)
F. nottii	PAS	EOW 1987-08	87-8-1, 87-8-2 to 87-8-6 (All *), 87-8-7 to 87-8-12 (All †), 1631-2*, 1631-3*, 1631-4, 1631-5 to 1631-7 (All *)	2 of 11 (18%)
F. nottii	MOB	EOW 1989-32	6654 to 6664, 6665*, 6666 to 6669	15 of 16 (94%)
F. escambiae	PER	EOW 1988-17‡	633-1 to 633-6 (All *) 6577 to 6580 (All *) 6581 to 6585	5 of 15 (33%)
F. escambiae	PER	EOW 1989-26‡	9-26-1, 9-26-2, 9-26-3*, 9-26-4 to 9-26-13, 9-26-15, 9-26-26, 9-26-19	15 of 16 (94%)
F. escambiae	ESC	EOW 1988-29	902-1 to 902-6 (All *)	0 of 6 (0%)
F. escambiae	ESC	EOW 1988-30	6616 to 6623, 6624*, 6625, 6626, 6627*, 6628 to 6631	14 of 16 (88%)
F. escambiae	BLA	EOW 1987-19	6558*, 6559 to 6568, 6569*, 6570 to 6573	14 of 16 (88%)
F. escambiae	BLA	EOW 1988-28	88-28-1 to 88-28-5 (All *) 88-28-6, 88-28-7, 88-28-8*, 6604 to 6614, 6615*	13 of 20 (65%)
F. escambiae	CHO	EOW 1988-25	525-1 to 525-4 (All *), 901-1 to 901-3 (All *), 901-4 to 901-9 (All †), 6586 to 6590, 6593 to 6595, 6596*, 6597 to 6603	15 of 23 (65%)

* Sample didn't work; † Sample not tried

Appendix 4. *Fundulus* specimens used in phylogenetic analysis of cytochrome *b* sequence. Tissue number is given, followed by number of specimens used with that tissue number, followed by unique tissue identifications, if applicable.

Fundulus nottii
KU T7501–7504, voucher KU 38027, Alabama R., Mobile Bay Drainage
KU T6654–6657, no voucher, Tombigbee R., Mobile Bay Drainage
Tissues EOW1987-08-1 and EOW1987-8-4 (not accessioned), Pascagoula River Drainage
KU T6638, no voucher, Pascagoula River Drainage
KU T6673, 6674, no voucher, Pascagoula River Drainage
KU T6689–6690, no voucher, Biloxi Bay Drainage
KU T6705, 6706, no voucher, Bay St. Louis Drainage
KU T6721–6723, no voucher, Pearl River Drainage

Fundulus escambiae
KU T6587, 6588, no voucher, Choctawhatchee River Drainage
KU T6558, 6559, no voucher, Blackwater River Drainage
KU T6604. 6605, no voucher, Blackwater River Drainage
KU T6581, 6582, no voucher, Perdido River Drainage
Tissue EOW1989-29-6 (not accessioned), Perdido River Drainage
KU T6616–6618, no voucher, Escambia River Drainage

Fundulus dispar – KU T7499, voucher KU 25197
Fundulus linealatus – KU T7497, 7498, voucher KU 25188

Literature Cited

Agassiz, L. 1854. Notice of a collection of fishes from the southern bend of the Tennessee River in the State of Alabama. American Journal of Science and Arts 2:353–369.

Araki, H., R. S. Waples, W. R. Ardren, B. Cooper, and M. S. Blouin. 2007. Effective population size of steelhead trout: influence of variance in reproductive success, hatchery programs, and genetic compensation between life-history forms. Molecular Ecology 16:953–966.

Arnqvist, G., M. Edvardsson, U. Friberg, and T. Nilsson. 2000. Sexual conflict promotes speciation in insects. Proceedings of the National Academy of Sciences of the USA 97:10460–10464.

Avise, J. C. 1989. Gene trees and organismal histories: A phylogenetic approach to population biology. Evolution 43:1192–1208.

Avise, J. C. 1994. Molecular Markers, Natural History, and Evolution. Chapman and Hall, New York.

Ayala, F. 2000. Neutralism and selectionism: the molecular clock. Gene 261:27–33.

Bagley, M. J., S. L. Anderson, and B. May. 2001. Choice of methodology for assessing genetic impacts of environmental stressors: polymorphism and reproducibility of RAPD and AFLP fingerprints. Ecotoxicology 10:239–244.

Barraclough, T. G., P. H. Harvey, and S. Nee. 1995. Sexual selection and taxonomic diversity in passerine birds. Proceedings of the Royal Society of London, Series B 259:211–215.

Baugh, T. A. 1981. Notes on the reproductibe behavior of 5 species of genus *Fundulus* in aquaria. Journal of Aquariculture 2:86–89.

Bellwood, D. R., L. van Herwerden, and N. Konow. 2004. Evolution and biogeography of marine angelfishes (Pisces: Pomacanthidae). Molecular Phylogenetics and Evolution 33:140–155.

Bensch, S., and M. Åkesson. 2005. Ten years of AFLP in ecology and evolution: why so few animals? Molecular Ecology 14:2899–2914.

Bernardi, G., G. Bucciarelli, D. Costagliola, D. R. Robertson, and J. B. Heiser. 2004. Evolutional of coral reef fish *Thalassoma* spp. (Labridae). 1. Molecular phylogeny and biogergraphy. Marine Biology 144:369–375.

Bollman, C. R. 1886. Notes on a collection of fishes from the Escambia River, with a description of a new species of *Zygonectes (Zygonectes escambiae)*. Proceedings of the United States National Museum 9:402–465.

Bolnick, D., and T. J. Near. 2005. Tempo of hybrid inviability in centrarchid fishes (Teleostei: Centrarchidae). Evolution 59:1754–1767.

Boschung, H. T., Jr., and R. L. Mayden. 2004. Fishes of Alabama. Smithsonian Books, Washington D.C.

Bowen, B. W., A. L. Bass, L. A. Rocha, W. S. Grant, and D. R. Robertson. 2001. Phylogeography of the trumperfishes (Aulostomus): Ring species complex on a global scale. Evolution 55:1029–1039.

Bromham, L. 2002. Molecular clocks in reptiles: Life history influences rate of molecular evolution. Molecular Biology and Evolution 19:302–309.

Busch, J. D., M. P. Miller, E. H. Paxton, M. K. Sogge, and P. Keim. 2000. Genetic variation in the endangered southwestern willow flycatcher. The Auk 117:586–595.

Cashner, R. C., J. S. Rogers, and J. M. Grady. 1992. Phylogenetic studies of the genus *Fundulus*. Pages 421–437 *in* R. L. Mayden, editor. Systematics, Historical Ecology, and North American Freshwater fishes. Stanford University Press, Stanford.

Congiu, L., I. Dupanloup, T. Patarnello, F. Fontana, R. Rossi, G. Arlati, and L. Zane. 2001. Identification of interspecific hybrids by amplified fragment length polymorphism. Molecular Ecology 10:2355–2359.

Coyne, J. A., and H. A. Orr. 1989. Patterns of speciation in *Drosophila*. Evolution 43:362–381.

Coyne, J. A., and H. A. Orr. 1997. "Patterns of speciation in *Drosophila"* revisited. Evolution 51:295–303.

Coyne, J. A., and H. A. Orr. 1998. The evolutionary genetics of speciation. Philosophical Transactions of the Royal Society of London Series B 353:287–305.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer, Sunderland, MA.

Craig, M. T., P. A. Hastings, and D. J. Pondella II. 2004. Speciation in the Central American Seaway: the importance of taxon sampling in the identification of trans-isthmian geminate pairs. Journal of Biogeography 31:1085–1091.

Crego, G. T., and M. S. Peterson. 1997. Salinity tolerance of four ecologically distinct species of *Fundulus* (Pisces: Fundulidae) from the northern Gulf of Mexico. Gulf of Mexico Science 15:45–49.

Domingues, V. S., G. Bucciarelli, V. C. Almada, and G. Bernardi. 2005. Historical colonization and demography of the Mediterranean damselfish, *Chromis chromis*. Molecular Ecology 14:4051–4063.

Donaldson, S. L., T. Chopin, and G. W. Saunders. 1998. Amplified fragment length polymorphism (AFLP) as a source of genetic markers for red algae. Journal of Applied Phycology 10:365–370.

Donaldson, S.-L., T. Chopin, and G.-W. Saunders. 2000. An assessment of the AFLP method for investigating population structure in the red alga *Chondrus crispus* Stackhouse (Gigartinales, Florideophyceae). Journal of Applied Phycology 12:25–35.

Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow. 1998. Rates of spontaneous mutation. Genetics 148:1667–1686.

Felsenstein, J. 2004. PHYLIP (Phylogenetic Inference Package). Distributed by the author, Department of Genome Sciences, University of Washington, Seattle, WA.

Foster, S. A. 1989. The implications of divergence in spatial nesting patterns in the geminate Caribbean and Pacific sergeant major damselfishes. Animal Behaviour 37:465–476.

Freeman, M. C. 1995. Movements by two small fishes in a large stream. Copeia 1995:361–367.

García-París, M., M. Alcobendas, D. Buckley, and D. B. Wake. 2003. Dispersal of viviparity across contact zones in Iberian populations of fire salamanders *(Salamandra)* inferred from discordance of genetic and morphological traits. Evolution 57:129–143.

Garman, S. 1895. The Cyprinodonts. Memoirs Of the Museum of Comparative Zoology 19:1–179.

Gaut, B. S., S. V. Muse, W. D. Clark, and M. T. Clegg. 1992. Relative rates of nucleotide substitution at the RBCL locus of monocotyledonous plants. Journal of Molecular Evolution 35:292–303.

Ghedotti, M. J., and M. J. Grose. 1997. Phylogenetic relationships of the *Fundulus nottii* species group (Fundulidae, Cyprinodontiformes) as inferred from the cytochrome *b* gene. Copeia 4:858–862.

Gillooly, J. F., A. P. Allen, G. B. West, and J. H. Brown. 2005. The rate of DNA evolution: Effects of body size and temperature on the molecular clock. Proceedings of the National Academy of Sciences of the USA 102:140–145.

Goulao, L., L. Monte-Corvo, and C.-M. Oliveira. 2001. Phenetic characterization of plum cultivars by high multiplex ratio markers: Amplified fragment length polymorphisms and inter-simple sequence repeats. Journal of the American Society for Horticultural Science 126:72–77.

Grant, W. S., and R. W. Leslie. 2001. Inter-ocean dispersal is an important mechanism in the zoogeography of hakes (Pisces: *Merluccius* spp.). Journal of Biogeography 28:699–721.

Graves, J. E., R. H. Rosenblatt, and G. N. Somero. 1983. Kinetic and electrophoretic differentiation of lactate dehydrogenases of teleost species pairs from the Atlantic and Pacific coasts of Panama. Evolution 37:30–37.

Gu, X., and W.-H. Li. 1992. Higher rates of amino acid substitution in rodents than in humans. Molecular Phylogenetics and Evolution 1:211–214.

Guiking, D., R. Lawson, U. Joger, and M. Wink. 2006. Evolution and phylogeny of the genus *Natrix* (Serpentes: Colubridae). Biological Journal of the Linnean Society 87:127–143.

Hartl, L., and S. Seefelder. 1997. Diversity of selected hop cultivars detected by fluorescent AFLP. Theoretical and Applied Genetics 96:112–116.

Heins, D. C., and J. R. Rooks. 1984. Life history of the naked sand darter, *Ammocrypta beanii,* in southeastern Mississippi. Pages 61–69 *in* D. G. Lindquist and L. M. Page, editors. Environmental Biology of Darters: Developments in Environmental Biology of Fishes, Volume 4. Dr. W. Junk Publishers, The Hague.

Heins, D. C. 1985. Life history traits fo the Florida sand darter *Ammocrypta bifascia,* and comparisons with the naked sand darter *Ammocrypta beanii*. The American Midland Naturalist 113:209–216.

Helfman, G. S., B. B. Collette, and D. E. Facey. 1997. The Diversity of Fishes. Blackwell Science, Inc., Malden, MA.

Holsinger, K. E., P. O. Lewis, and D. K. Dey. 2002. A Bayesian approach to inferring population structure from dominant markers. Molecular Ecology 11:1157–1164.

Hongtrakul, V., G. M. Huestis, and S. J. Knapp. 1997. Amplified fragment length polymorphisms as a tool for DNA fingerprinting sunflower germplasm: genetic diversity among oilseed inbred lines. Theoretical and Applied Genetics 95:400–407.

Iba, Y., and S.-I. Sano. 2007. MidCretaceous step-wise demise of the carbonate platform biota in the Northwest Pacific and establishment of the North Pacific Biotic Province. Palaeogeography, Palaeoclimatology, Palaeoecology 246:488–501.

Johnston, C. E. 1989. Spawning in the eastern sand darter, *Ammocrypta pellucida* (Pisces: Percidae), with comments on the phylogeny of *Ammocrypta* and related taxa. Transactions of the Illinois Academy of Science 82:163–168.

Jordan, D. S., and B. W. Evermann. 1896. The fishes of North and Middle America: a descriptive catalogue of the species of fish-like vertebrates found in the waters of North America, north of the Isthmus of Panama. Bulletin of the U.S. National Museum 47:1–3313.

Kruse, I., T. B. H. Reusch, and M. V. Schneider. 2003. Sibling species or poecilogony in the polychaete *Scoloplos armiger?* Marine Biology 142:937–947.

Lee, D. S., C. R. Gilbert, C. H. Hocutt, R. E. Jenkins, D. E. McAllister, and J. R. Stauffer. 1980. Atlas of North American Freshwater Fishes. North Carolina State Museum of Natural History.

Li, W.-H., D. L. Ellesworth, J. Krushkal, B. H.-J. Chang, and D. Hewett-Emmett. 1996. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. Molecular Phylogenetics and Evolution 5:182–187.

Li, W.-H., and M. Tanimura. 1987. The molecular clock runs more slowly in man tghan in apes and monkeys. Nature 326:93–96.

Lijtmaer, D. A., B. Mahler, and P. L. Tubaro. 2003. Hybridization and postzygotic isolation patterns in pigeons and doves. Evolution 57:1411–1418.

Lynch, M., and B. G. Milligan. 1994. Analysis of population genetic structure with RAPD markers. Molecular Ecology 3:91–99.

Martin, A. P., and S. R. Palumbi. 1993. Body size, metabolic rate, generation time, and the molecular clock. Proceedings of the National Academy of Sciences of the USA 90:4087–4091.

Mendelson, T. C. 2003. Sexual isolation evolves faster than hybrid inviability in a diverse and sexually dimorphic genus of fish (Percidae: *Etheostoma)*. Evolution 57:317–327.

Michel, A. P., M. J. Ingrasci, B. J. Schemerhorn, M. Kern, G. Le Goff, M. Coetzee, N. Elissa, D. Fontenille, J. Vulule, T. Lehmann, N. F. Sagnon, C. Costantini, and N. J. Besansky. 2005. Rangewide population genetic structure of the African malaria vector *Anopheles funestus*. Molecular Ecology 14:4235–4248.

Miller, M. P. 1997. Tools for population genetic analysis (TFPGA) 1.3: a Windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by the author at http://bioweb.usu.edu/mpmbio/

Miller, M. P., D. W. Blinn, and P. Keim. 2002. Correlations between observed dispersal capabilities and patterns of genetic differentiation in populations of four aquatic insect species from the Arizona White Mountains, U.S.A. Freshwater Biology 47:1660–1673.

Mitra, S., H. Landel, and S. Pruett-Jones. 1996. Species richness covaries with mating system in birds. Auk 113:544–551.

Mock, K. E., T. C. Theimer, R. O. E., D. L. Greenberg, and P. Keim. 2002. Genetic variation across the historical range of the wild turkey *(Meleagris gallopavo)*. Molecular Ecology 11:643–657.

Murphy, R. W., T. L. Trepanier, and D. J. Morafka. 2006. Conservation genetics, evolution and distinct population segments of the Mojave fringe-toed lizard, *Uma scoparia*. Journal of Arid Environments 67 (Supplement S):226–247.

Muss, A., D. R. Robertson, C. A. Stepien, P. Wirz, and B. W. Bowen. 2001. Phylogeography of *Ophioblennius*: The role of ocean currents and geography in reef fish evolution. Evolution 55:561–572.

Near, T. J., J. C. Porterfield, and L. M. Page. 2000. Evolution of cytochrome *b* and the molecular systematics of *Ammocrypta* (Percidae: Etheostomatinae). Copeia 2000:701–711.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 23:341–369.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Nylander, J. A. A. 2004. MrModeltest v2. Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden.

Osinov, A. G., and V. S. Lebedev. 2000. Genetic divergence and phylogeny of the Salmononae based on allozyme data. Journal of Fish Biology 57:354–381.

Palacios, C., S. Kresovich, and F. Gonzalez-Candelas. 1999. A population genetic study of the endangered plant species *Limonium dufourii* (Plumbaginaceae) based on amplified fragment length polymorphism (AFLP). Molecular Ecology 8:645–657.

Porter, A. H. 1990. Testing nominal species boundaries using gene flow statistics: the taxonomy of two hybridizing admiral butterflies (Limenitis: Nymphalidae). Systematic Zoology 39:131–147.

Poulkakis, N., P. Lymberakis, E. Valakos, P. Pafilis, E. Zouros, and M. Mylonas. 2005. Phylogeography of Balkan wall lizard *(Podarcis taurica)* and its relatives inferred from mitochondrial DNA sequences. Molecular Ecology 14:2433–2443.

Presgraves, D. C. 2002. Patterns of postzygotic isolation in Lepidoptera. Evolution 56:1168–1183.

Price, T. D., and M. M. Bouvier. 2002. The evolution of F1 post-zygotic incompatibilities in birds. Evolution 56:2083–2089.

Price, R. C., and K. N. Whetsone. 1977. Lateral stream migration as evidence for regional geologic structures in the eastern Gulf Coastal Plane. Southeastern Geologist 18:129–147.

Quinteiro, J., R. Vidal, and M. Rey-Mendez. 2000. Phylogeny and biogeographic history of hake (genus *Merluccius*), inferred from mitochondrial DNA control-region sequences. Marine Biology 136:163–174.

Ronquist, F., and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Ross, S. T., and J. A. Baker. 1983. The response of fishes to periodic spring floods in a southeastern stream. The American Midland Naturalist 109:1–14.

Ross, S. T., and W. M. Brenneman. 2001. Inland Fishes of Mississippi. University Press of Mississippi.

Russell, S. T. 2003. Evolution of intrinsic post-zygotic reproductive isolation in fish. Annales Zoologici Fennici 40:321–329.

Ruzzante, D. E., S. J. Walde, V. E. Cussac, D. M. L., J. Seibert, S. Ortubay, and E. Habit. 2006. Phylogeography of the Percichthyidae (Pisces) in Patagonia: roles of orogeny, glaciation, and volcanism. Molecular Ecology 15:2949–2968.

Sasa, M., P. T. Chippendale, and N. A. Johnson. 1998. Patterns of postzygotic isolation in frogs. Evolution 52:1811–1820.

Shaw, K. A., A. M. Simons, and E. O. Wiley. 1999. A reexamination of the phylogenetic relationships of the sand darters (Teleostei: Percidae). Scientific Papers of the Natural History Museum of The University of Kansas 12:1–6.

Simons, A. M. 1992. Phylogenetic relationships of the *Boleosoma* species group (Percidae: *Etheostoma)*. Pages 268–292 *in* R. L. Mayden, editor. Systematics, historical ecology, and North American freshwater fishes. Stanford University Press, Stanford, CA.

Slatkin, M. 1985. Rare alleles as indicators of gene flow. Evolution 39:53–65.

Smith, G. R. 1981. Late Cenozoic freshwater fishes of North America. Annual Review of Ecology and Systematics 12:163–193.

Song, C. B., T. J. Near, and L. M. Page. 1998. Phylogenetic relations among percid fishes as inferred from mitochondrial cytochrome *b* DNA sequence data. Molecular Phylogenetics and Evolution 10:343–353.

Swift, C. C., C. R. Gilbert, S. A. Bortone, G. H. Burgess, and R. W. Yerger. 1986. Zoogeography of the freshwater fishes of the southeastern United States: Savannah River to Lake Pontchartrain. Pages 213–257 *in* C. H. Hocutt and E. O. Wiley, editors. The Zoogeography of North American Freshwater Fishes. John Wiley and Sons, New York.

Swofford, D. L. 2003. PAUP*: Phylogenetic analysis using parsimony (* and other methods). Sinauer Associates, Sunderland, Massachusetts.

Taylor, C. A., and B. M. Burr. 1996. Reproductive biology of the northern starhead topminnow, *Fundulus dispar* (Osteichthyes: Fundulidae), with a review of data for freshwater members of the genus. American Midland Naturalist 137:151–164.

Thorpe, R. S., D. L. Leadbeater, and C. E. Pook. 2005. Molecular clocks and geological dates: cytochrome *b* of *Anolis extremus* substantially contradicts dating of Barbados emergence. Molecular Ecology 14:2087–2096.

Tohme, J., D. O. Gonzalez, S. Beebe, and M. C. Duque. 1996. AFLP analysis of gene pools of wild bean core collection. Crop Science 36:1375–1384.

Vawter, A. T., R. H. Rosenblatt, and G. C. Gorman. 1980. Genetic divergence among fishes of the eastern Pacific and the Caribbean: support for the molecular clock. Evolution 34:705–711.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Van De Leet, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23:4407–4414.

Weir, B. S. 1996. Genetic Data Analysis II. Sinauer Associates, Inc., Sunderland, MA.

Wellington, G. M., and D. R. Robertson. 2001. Variation in larval life-history traits among reef fishes across the Isthmus of Panama. Marine Biology 138:11–22.

Whitehead, A., S. L. Anderson, K. M. Kuivila, J. L. Roach, and B. May. 2003. Genetic variation among interconnected populations of *Catostomus occidentalis*: Implications for distinguishing impacts of contaminants from biogeographical structuring. Molecular Ecology 12:2817–2833.

Wiley, E. O. 1977. The phylogeny and systematics of the *Fundulus nottii* species group (Teleostei: Cyprinodontidae). Occasional Papers of the Museum of Natural History of the University of Kansas 66:1–31.

Wiley, E. O. 1981. Phylogenetics: The theory and practice of phylogenetic systematics. John Wiley and Sons, New York.

Wiley, E. O., and R. H. Hagen. 1997. Mitochondrial DNA sequence variation among the sand darters (Percidae: Teleostei). Pages 75–96 *in* T. D. Kocher and C. A. Stepien, editors. Molecular Systematics of Fishes. Academic Press, San Diego.

Wiley, E. O., and R. L. Mayden. 1985. Species and speciation in phylogenetic systematics, with examples from the North American fish fauna. Annals of the Missouri Botanical Gardens 72:596–635.

Williams, J. D. 1975. Systematics of the percid fishes of the subgenus *Ammocrypta,* genus *Ammocrypta,* with descriptions of two new species. Bulletin of the Alabama Museum of Natural History 1:1–56.

Wilson, C. C., and L. Bernachez. 1998. The ghost of hybrids past: fixation of arctic charr *(Salvelinus alpinus)* mitochondrial DNA in an introgressed population of lake trout *(S. namaycush)*. Molecular Ecology 7:127–132.

Wright, S. 1978. Evolution and the Genetics of Populations Volume 4. Variability within and among Natural Populations. University of Chicago Press, Chicago.

Wu, C.-I., and W.-H. Li. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proceedings of the National Academy of Sciences of the USA 82:1741–1745.

Yang, Z., and R. Nielsen. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. Journal of Molecular Evolution 46:409–418.

Yeh, F. C., R.-C. Yang, and T. Boyle. 1997. Popgene; Microsoft Window-based Freeware for Population Genetic Analysis. http://www.ualberta.ca/~fyeh/index.htm